PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/10, 15/62, C07K 14/01, C12N 1/00, 15/86 | A2

(11) International Publication Number: **WO 00/06717**

(43) International Publication Date: 10 February 2000 (10.02.00)

(21) International Application Number: PCT/US99/16596

(22) International Filing Date: 22 July 1999 (22.07.99)

(30) Priority Data:

60/094,291	27 July 1998 (27.07.98)	US
60/103,514	8 October 1998 (08.10.98)	US
60/133,296	10 May 1999 (10.05.99)	US
60/134,870	19 May 1999 (19.05.99)	US

(71) Applicant *(for all designated States except US)*: GENENTECH, INC. [US/US]; 1 DNA Way, South San Francisco, CA 94080–4990 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SIDHU, Sachdev, S. [CA/US]; Apartment 203, 574 – 3rd Street, San Francisco, CA 94107 (US). WEISS, Gregory, A. [US/US]; 733 Fairfield Road #3, Burlingame, CA 94010 (US). WELLS, James, A. [US/US]; 1342 Columbus Avenue, Burlingame, CA 94010 (US).

(74) Agents: SCHWARTZ, Timothy, R. et al.; Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080–4990 (US).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: IMPROVED TRANSFORMATION EFFICIENCY IN PHAGE DISPLAY THROUGH MODIFICATION OF A COAT PROTEIN

(57) Abstract

The transformation yield of electroporation is increased by using higher DNA concentrations and DNA affinity purification. Fusion proteins of a viral coat protein variant and a heterologous polypeptide are useful in phage display systems.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania
AM	Armenia
AT	Austria
AU	Australia
AZ	Azerbaijan
BA	Bosnia and Herzegovina
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
BY	Belarus
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CN	China
CU	Cuba
CZ	Czech Republic
DE	Germany
DK	Denmark
EE	Estonia
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GE	Georgia
GH	Ghana
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IL	Israel
IS	Iceland
IT	Italy
JP	Japan
KE	Kenya
KG	Kyrgyzstan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
KZ	Kazakstan
LC	Saint Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Lithuania
LU	Luxembourg
LV	Latvia
MC	Monaco
MD	Republic of Moldova
MG	Madagascar
MK	The former Yugoslav Republic of Macedonia
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
MX	Mexico
NE	Niger
NL	Netherlands
NO	Norway
NZ	New Zealand
PL	Poland
PT	Portugal
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SG	Singapore
SI	Slovenia
SK	Slovakia
SN	Senegal
SZ	Swaziland
TD	Chad
TG	Togo
TJ	Tajikistan
TM	Turkmenistan
TR	Turkey
TT	Trinidad and Tobago
UA	Ukraine
UG	Uganda
US	United States of America
UZ	Uzbekistan
VN	Viet Nam
YU	Yugoslavia
ZW	Zimbabwe

IMPROVED TRANSFORMATION EFFICIENCY IN PHAGE DISPLAY THROUGH MODIFICATION OF A COAT PROTEIN

FIELD OF THE INVENTION

The invention relates to fusion proteins of a polypeptide and a coat protein of a virus, where a coat protein is not a wild type coat protein. The invention also relates to replicable expression vectors which contain a gene encoding the fusion protein, host cells containing the expression vectors, a virus which displays the fusion protein on the surface of the virus, libraries of the virus displaying a plurality of different fusion proteins on viral surfaces and methods of using these compositions.

The invention also relates to a method of transforming cells by electroporating cells to improve transformation efficiency. In various preferred embodiments, the transformation is performed in the presence of a high concentration of DNA; in the presence of a high concentration of cells; with highly purified DNA; with specific host cells; or with combinations of these. When used to prepare libraries, for example phage display libraries, these improvements allow for the construction of larger libraries in a single electroporation step than has been previously possible. The invention is also directed to a method of producing a product polypeptide by transforming host cells using the method of the invention.

DISCUSSION OF THE BACKGROUND

Bacteriophage (phage) display is a technique by which variant polypeptides are displayed as fusion proteins to the coat protein on the surface of bacteriophage particles (Scott, J.K. and Smith, G. P. (1990) *Science* 249: 386). The utility of phage display lies in the fact that large libraries of selectively randomized protein variants (or randomly cloned cDNAs) can be rapidly and efficiently sorted for those sequences that bind to a target molecule with high affinity. Display of peptide (Cwirla, S. E. et al. (1990) *Proc. Natl. Acad. Sci. USA*, 87:6378) or protein (Lowman, H.B. et al. (1991) *Biochemistry*, 30:10832; Clackson, T. et al. (1991) *Nature*, 352: 624; Marks, J. D. et al. (1991), *J. Mol. Biol.*, 222:581; Kang, A.S. et al. (1991) *Proc. Natl. Acad. Sci. USA*, 88:8363) libraries on phage have been used for screening millions of polypeptides for ones with specific binding properties (Smith, G. P. (1991) *Current Opin. Biotechnol.*, 2:668). Sorting phage libraries of random mutants requires a strategy for constructing and propagating a large number of variants, a procedure for affinity purification using the target receptor, and a means of evaluating the results of binding enrichments. U.S. 5,223,409; U.S. 5,403,484; U.S. 5,571,689; U.S. 5,663,143.

Typically, variant polypeptides are fused to a gene III protein, which is displayed at one end of the viron. Alternatively, the variant polypeptides may be fused to the gene VIII protein, which is the major coat protein of the viron. Such polyvalent display libraries are constructed by replacing the phage gene III with a cDNA encoding the foreign sequence fused to the amino

terminus of the gene III protein. This can complicate efforts to sort high affinity variants from libraries because of the avidity effect; phage can bind to the target through multiple point attachment. Moreover, because the gene III protein is required for attachment and propagation of phage in the host cell, e.g., *E. coli*, the fusion protein can dramatically reduce infectivity of the progeny phage particles.

To overcome these difficulties, monovalent phage display was developed in which a protein or peptide sequence is fused to a portion of a gene III protein and expressed at low levels in the presence of wild-type gene III protein so that particles display mostly wild-type gene III protein and one copy or none of the fusion protein (Bass, S. et al. (1990) *Proteins*, 8:309; Lowman, H.B. and Wells, J.A. (1991) *Methods: a Companion to Methods in Enzymology*, 3:205). Monovalent display has advantages over polyvalent phage display in that progeny phagemid particles retain full infectivity. Avidity effects are reduced so that sorting is on the basis of intrinsic ligand affinity, and phagemid vectors, which simplify DNA manipulations, are used. See also U.S. 5,750,373 and U.S. 5,780,279. Others have also used phagemids to display proteins, particularly antibodies. U.S. 5,667,988; U.S. 5,759,817; U.S. 5,770,356; and U.S. 5,658,727.

A two-step approach has been used to select high affinity ligands from peptide libraries displayed on M13 phage. Low affinity leads were first selected from naive, polyvalent libraries displayed on the major coat protein (protein VIII). The low affinity selectants were subsequently transferred to the gene III minor coat protein and matured to high affinity in a monovalent format. Unfortunately, extension of this methodology from peptides to proteins has been difficult. Display levels on protein VIII vary with fusion length and sequence. Increasing fusion size generally decreases display. Thus, while monovalent phage display has been used to affinity mature many different proteins, polyvalent display on protein VIII has not been applicable to most protein scaffolds.

Although most phage display methods have used filamentous phage, lambdoid phage display systems (WO 95/34683; U.S. 5,627,024), T4 phage display systems (Ren, Z-J. et al. (1998) Gene 215:439; Zhu, Z. (1997) CAN 33:534; Jiang, J. et al. (1997) can 128:44380; Ren, Z-J. et al. (1997) CAN 127:215644; Ren, Z-J. (1996) Protein Sci. 5:1833; Efimov, V. P. et al. (1995) Virus Genes 10:173) and T7 phage display systems (Smith, G. P. and Scott, J.K. (1993) Methods in Enzymology, 217, 228-257; U.S. 5,766,905) are also known.

Many other improvements and variations of the basic phage display concept have now been developed. These improvements enhance the ability of display systems to screen peptide libraries for binding to selected target molecules and to display functional proteins with the potential of screening these proteins for desired properties. Combinatorial reaction devices for phage display reactions have been developed (WO 98/14277) and phage display libraries have been used to analyze and control bimolecular interactions (WO 98/20169; WO 98/20159) and properties of constrained helical peptides (WO 98/20036). WO 97/35196 describes a method of isolating an affinity ligand in which a phage display library is contacted with one solution in which

the ligand will bind to a target molecule and a second solution in which the affinity ligand will not bind to the target molecule, to selectively isolate binding ligands. WO 97/46251 describes a method of biopanning a random phage display library with an affinity purified antibody and then isolating binding phage, followed by a micropanning process using microplate wells to isolate high affinity binding phage. The use of *Staphlylococcus aureus* protein A as an affinity tag has also been reported (Li et al. (1998) Mol Biotech., 9:187). WO 97/47314 describes the use of substrate subtraction libraries to distinguish enzyme specificities using a combinatorial library which may be a phage display library. A method for selecting enzymes suitable for use in detergents using phage display is described in WO 97/09446. Additional methods of selecting specific binding proteins are described in U.S. 5,498,538; U.S. 5,432,018; and WO 98/15833.

Methods of generating peptide libraries and screening these libraries are also disclosed in U.S. 5,723,286; U.S. 5,432,018; U.S. 5,580,717; U.S. 5,427,908; and U.S. 5,498,530. See also U.S. 5,770,434; U.S. 5,734,018; U.S. 5,698,426; U.S.5,763,192; and U.S. 5,723,323.

Methods which alter the infectivity of phage are also known. WO 95/34648 and U.S. 5,516,637 describe a method of displaying a target protein as a fusion protein with a pilin protein of a host cell, where the pilin protein is preferably a receptor for a display phage. U.S. 5,712,089 describes infecting a bacteria with a phagemid expressing a ligand and then superinfecting the bacteria with helper phage containing wild type protein III but not a gene encoding protein III followed by addition of a protein III-second ligand where the second ligand binds to the first ligand displayed on the phage produced. See also WO 96/22393. A selectively infective phage system using non-infectious phage and an infectivity mediating complex is also known (U.S. 5,514,548).

Phage systems displaying a ligand have also been used to detect the presence of a polypeptide binding to the ligand in a sample (WO/9744491), and in an animal (U.S. 5,622,699). Methods of gene therapy (WO 98/05344) and drug delivery (WO 97/12048) have also been proposed using phage which selectively bind to the surface of a mammalian cell.

Further improvements have enabled the phage display system to express antibodies and antibody fragments on a bacteriophage surface, allowing for selection of specific properties, i.e., binding with specific ligands (EP 844306; U.S. 5,702,892; U.S. 5,658,727) and recombination of antibody polypeptide chains (WO 97/09436). A method to generate antibodies recognizing specific peptide - MHC complexes has also been developed (WO 97/02342). See also U.S. 5,723,287; U.S. 5,565,332; and U.S. 5,733,743.

U.S. 5,534,257 describes an expression system in which foreign epitopes up to about 30 residues are incorporated into a capsid protein of a MS-2 phage. This phage is able to express the chimeric protein in a suitable bacterial host to yield empty phage particles free of phage RNA and other nucleic acid contaminants. The empty phage are useful as vaccines.

The degree of expression of polypeptides as fusion proteins on the surface of bacteriophage particles is variable and depends, to some extent, on the size of the polypeptide. Conventional phage display systems use wild type phage coat proteins and fuse the heterologeous

polypeptide to the amino terminus of the wild type amino acid sequence or an amino terminus resulting from truncation of the wild type coat protein sequence. Segments of linker amino acids have also been added to the amino terminus of the wild type coat protein sequence to improve selection and target binding.

Notwithstanding numerous modifications and improvements in phage technology, a need continues to exist for improved methods of displaying polypeptides as fusion proteins in phage display methods.

Methods of transforming cells to introduce new DNA are of great practical interest in molecular biology and modern genetic engineering. Early methods involved chemical treatment of bacteria with solutions of metal ions, generally calcium chloride, followed by heating to produce competent bacteria capable of functioning as recipient bacteria and able to take up heterologous DNA derived from a variety of sources. These early protocols provided transformation yields of about 10^5 - 10^6 transformed colonies per µgram of plasmid DNA. Subsequent improvements using different cations, longer treatment times and other chemical agents have allowed improvements in transformation efficiency of up to about 10^8 colonies/µgram of DNA. Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd edition, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, page 1.74.

Cells can also be transformed using high-voltage electroporation. Electroporation is suitable to introduce DNA into eukaryotic cells (e.g. animal cells, plant cells, etc.) as well as bacteria, e.g., *E. coli*. Sambrook et al., *ibid*, pages 1.75, 16.54-16.55. Different cell types require different conditions for optimal electroporation and preliminary experiments are generally conducted to find acceptable levels of expression or transformation. For mammalian cells, voltages of 250-750 V/cm result in 20-50% cell survival. An electric pulse length of 20-100 ms at a temperature ranging from room temperature to 0°C and below using a DNA concentration of 1-40 µgram/mL are typical parameters. Transfection efficiency is reported to be higher using linear DNA and when the cells are suspended in buffered salt solutions than when suspended in nonionic solutions. Sambrook et al., above, pages 16.54-16.55.

Dower et al., 1988, *Nucleic Acids Research*, 16:6127-6145 extensively studied high-efficiency transformation of *E. coli* by high-voltage electroporation. This study evaluated numerous parameters, including electrical variables such as the effect of field strength and pulse length, the effect of DNA concentration and cell concentration on the recovery of transformants, accuracy, reproducibility, etc. and provided a protocol for high-efficiency electrotransformation of *E. coli* cells. The optimized protocol of Dower et al. uses cells concentrated in a range of at least 2 and up to 4 x 10^{10}/mL, a DNA concentration of from about 1 to 10 µgrams/mL, 12.5-16.7 kV/cm, 3-25 µF and the electroporation is conducted at 0°C (ice temperature). These studies were conducted with highly purified closed circular plasmid DNA, which is known to give high transformation efficiencies. Dower et al. report transformation efficiencies of 10^9 - 10^{10}

transformants/μgram of DNA achieved by highly optimizing these parameters. For library formation, Dower et al. suggests using a DNA concentration of less than 10 nanogram/mL and a cell concentration of greater than 3 x 10^{10} to minimize co-transformants. See also U.S. 4,910,140 and U.S. 5,186,800 to Dower et al. and U.S. 4,849,355 to Wong.

Several attempts have been made to improve the design of electroporation apparatus (see, for example, U.S. 5,173,158; U.S. 5,098,843; U.S. 5,422,272; U.S. 5,232,856; and U.S. 5,283,194) and to improve electroporation of specific cells (see U.S. 5,128,257). U.S. 5,124,259 describes an improved buffer for electroporation. U.S. 4,956,288 describes a method for producing cells containing foreign DNA in high copy numbers.

The attainment of higher transformation efficiencies by optimizing the electroporation parameters has been difficult. The use of higher voltages and longer pulses results in an increase in cell death, decreasing the total number of transformed cells. Highly optimized electroporation still results in about 50-75% cell death. Dower et al. represents an important investigation of the parameters of electroporation and the protocol described in this paper has formed the basis of more recent electroporation procedures

An important emerging use of cell transformations, including electroporation, is the preparation of peptide and protein variant libraries. In these applications, a replicable transcription vector, for example a plasmid, is reacted with a restriction enzyme to open the plasmid DNA, desired coding DNA is ligated into the plasmid to form a library of vectors each encoding a different variant, and cells are transformed with the library of transformation vectors in order to prepare a library of polypeptide variants differing in amino acid sequence at one or more residues. The library of peptides can then be selectively panned for peptides which have or do not have particular properties. A common property is the ability of the variant peptides to bind to a cell surface receptor, an antibody, a ligand or other binding partner, which may be bound to a solid support. Variants may also be selected for their ability to catalyze specific reactions, to inhibit reactions, to inhibit enzymes, etc.

In one application, bacteriophage (phage), such as filamentous phage, are used to create phage display libraries by transforming host cells with phage vector DNA encoding a library of peptide variants. J.K. Scott and G.P. Smith, *Science,* (1990), 249:386-390. Phagemid vectors may also be used for phage display. Lowman and Wells, 1991, Methods: A Companion to Methods in Enzymology, 3:205-216. The preparation of phage and phagemid display libraries of peptides and proteins, e.g. antibodies, is now well known in the art. These methods generally require transforming cells with phage or phagemid vector DNA to propagate the libraries as phage particles having one or more copies of the variant peptides or proteins displayed on the surface of the phage particles. See, for example, Barbas et al., *Proc. Natl. Acad. Sci., USA,* (1991), 88:7978-7982; Marks et al., *J. Mol. Biol.*, (1991), 222:581-597; Hoogenboom and Winter, *J. Mol. Biol.*, (1992), 227:381-388; Barbas et al., *Proc. Natl. Acad. Sci., USA,* (1992), 89:4457-4461; Griffiths et al., *EMBO Journal,* (1994), 13:3245-3260; de Kruif et al., *J. Mol. Biol.*, (1995), 248:97-105;

Bonnycastle et al., *J. Mol. Biol.*, (1996), 258:747-762; and Vaughan et al., *Nature Biotechnology* (1996), 14:309-314. The library DNA is prepared using restriction and ligation enzymes in one of several well known mutagenesis procedures, for example, cassette mutagenesis or oligonucleotide-mediated mutagenesis.

A recurring problem with transformation by electroporation, in particular with phage or phagemid vector DNA libraries, is the low transformation efficiency which has generally been in the range of 10^7 - 10^8 transformations/μgram of DNA. The low transformation efficiency has limited the size of libraries which can be prepared with a single electroporation step. Vaughan et al., above, describe a modified procedure in which several hundred electroporations were conducted to achieve a library with about 10^{10} recombinants.

Reaction mixtures obtained by enzymatic manipulation of DNA and RNA contain proteins, salts, etc., which are contaminants of the desired DNA or RNA. To obtain the purified nucleic acid, these mixtures are usually extracted with phenol/chloroform or similar solvent and then the DNA is precipitated with ethanol and resuspended in an appropriate amount of water or buffer to provide the DNA concentrations recommended by Dower et al. Bonnycastle et al., above, describe extracting a ligation reaction with chloroform/phenol/isoamyl alcohol followed by resuspension of the DNA in water and desalting by filtration over an exclusion membrane. This procedure allowed electroporation of electrocompetent MC1061 *E. coli* cells using a DNA concentration of about 20 μgrams/mL.

Despite two decades of research into electroporation and parameters affecting transformation efficiency, a need continues to exist for improved electroporation processes, in particular, for the transformation of cells with libraries of phage and phagemid DNA vectors.

SUMMARY OF THE INVENTION

Conventional phage display methods use wild type coat protein sequences, presumably to enhance stability of the phage particles and to increase the frequency of incorporation of fusion proteins into the coat of stable phage particles. It has now been discovered that stable viral particles can be prepared which incorporate fusion proteins containing a heterologeous polypeptide of interest fused to a coat protein of the virus, preferably a major coat protein, where the coat protein is not a wild type coat protein of the virus, that is, where the coat protein has one or more amino acid substitutions, deletions or additions. This result is unexpected since prior phage display techniques have utilized wild type coat protein sequences and the incorporation of heterologeous polypeptides as fusion proteins into phage particles is expected to have a generally deleterious effect on normal phage packaging and phage particle production.

One object of the present invention is to provide a fusion protein containing a heterologeous polypeptide fused to a coat protein of a virus where the coat protein does not have the wild type coat protein sequence. A further object is to provide a replicable expression vector

containing a gene fusion which encodes this fusion protein. A further object is to provide host cells containing the replicable expression vector.

Another object of the invention is to provide a library of the replicable expression vectors where the vectors contain a plurality of different gene fusions encoding a plurality of variant fusion proteins. A further object is to provide a library of virus particles which display a plurality of variant fusion proteins where the fusion proteins contain a coat protein variant which does not have the wild type coat protein sequence, and the host cells containing the vector libraries.

An additional object of the invention is to provide a method of constructing the libraries of expression vectors and virus particles.

A further object is to provide a method of modulating the number of fusion proteins which are displayed on the surface of a phage or phagemid particle through the use of the fusion protein of the invention.

Another object of the present invention is to provide an improved method of transforming cells by electroporating competent cells in the presence of heterologous DNA.

A further object of the invention is to provide an improved strain of *E. coli* cells which have improved characteristics and allow higher transformation yields with electroporation.

A further object is to provide a method of producing a product polypeptide by culturing a host cell transformed with a replicable expression vector where the host cells have been transformed using the method of the present invention, and product polypeptides produced by this process.

One embodiment of the invention is a method, which includes constructing a library containing a plurality of replicable expression vectors, each expression vector containing a transcription regulatory element operably linked to a gene fusion encoding a fusion protein, where the gene fusion contains a first gene encoding a first polypeptide and a second gene encoding a phage major coat protein, where the library contains a plurality of second genes encoding variant phage major coat proteins. The method may further include transforming suitable host cells with the library of vectors and culturing the transformed cells under conditions suitable to form the fusion proteins. Preferably, the vector is phage or phagemid DNA and the culturing is sufficient to form phage or phagemid particles which display fusion proteins on the surfaces thereof. The method may also include contacting the phage or phagemid particles with a target molecule so that at least a portion of the particles bind to the target molecule , and separating the particles that bind from those that do not bind. The method may further include selecting a bound particle, constructing a second library containing a plurality of replicable expression vectors, each expression vector containing a transcription regulatory element operably linked to a second gene fusion encoding a second fusion protein, where the second gene fusion contains a third gene encoding a second polypeptide and a fourth gene encoding the major coat protein variant of the fusion protein displayed on the surface of the selected bound particle, where the library contains a

plurality of third genes encoding variant second polypeptides. The first polypeptide and the third polypeptide may be the same or different.

In another embodiment, the invention is a method including the steps:

(a) constructing a first library containing a plurality of first replicable expression vectors, each expression vector comprising a transcription regulatory element operably linked to a first gene fusion encoding a first fusion protein, wherein the first gene fusion comprises

a first gene encoding a first polypeptide and

a second gene encoding a phage major coat protein, and

wherein the first library contains a plurality of first vectors encoding second genes encoding variant phage major coat proteins;

(b) transforming suitable host cells with the first library of vectors and culturing the transformed cells under conditions suitable to form phage or phagemid particles;

(c) contacting the phage or phagemid particles with a target molecule so that at least a portion of the particles bind to the target molecule;

(d) separating particles that bind from those that do not bind;

(e) selecting a particle;

(f) constructing a second replicable expression vector comprising a transcription regulatory element operably linked to a second gene fusion encoding a second fusion protein, wherein the second gene fusion comprises

a third gene encoding a second polypeptide and

a fourth gene encoding the major coat protein variant of the fusion protein displayed on the surface of the selected bound particle;

(g) transforming suitable host cells with the second vector, culturing the transformed cells under conditions suitable to form phage or phagemid particles displaying the second fusion protein on the surface thereof; and

(h) separating particles that bind from those that do not bind.

The method may also include constructing a second library of second vectors containing a plurality of third genes encoding variant second polypeptides. The method may further include steps:

(i) selecting a particle;

(j) constructing a third expression vector comprising a transcription regulatory element operably linked to the third gene of the particle selected in step (i); and

(k) transforming suitable host cells with the third vector obtained in step (j) and culturing the transformed cells under conditions suitable for forming the second polypeptide.

These and other objects which will become apparent in the course of the following descriptions of exemplary embodiments have been achieved by the present method of transforming cells by electroporating competent cells in the presence of heterologous DNA, where the DNA is purified by affinity purification, and is preferably present at a concentration of about 1

picogram/mL to about 500 μgram/mL. The DNA is generally present at a concentration of a few to several hundred nanograms/mL or greater, preferably about 1 to about 50 μgrams/mL or greater, even more preferably about 70 μgrams/mL or greater, and may be present at a concentration of greater than 100 μgrams/mL to about 500 μgrams/mL.

In part, the present invention is also based on the discovery that prior art methods of preparing DNA for electroporation, for example the preparation of clonable recombinant DNA, using phenol extractions and ethanol precipitation, have generally resulted in DNA solutions having unacceptably high conductance. Electroporation instruments are generally configured to have a sample cell in parallel with a capacitor and a resistor (R2) to control the electric pulse duration through the sample. Ideally, the resistance of the sample (R1) should be much greater than that of R2 so that the electric pulse decays mainly through R2. In a preferred electroporation, where essentially the entire discharge occurs through R2, the time constant would approach the theoretical maximum where R1 is infinite. DNA is an ionic molecule, and thus DNA electroporation samples have an inherent conductance. Furthermore, DNA preparations containing electrically charged impurities such as proteins, salts, buffers, etc., introduce additional conductance. The volume of a DNA preparation (and thus the mass of DNA) which can be introduced into an electroporation reaction is limited by the conductance of the preparation. As the conductance of the sample increases, R1 decreases and becomes significant in comparison to R2, i.e. a significant proportion of the electric pulse is discharged through R1. This results in a decrease in the time constant and a decrease in the transformation efficiency. Further increases in sample conductance result in electric arcing across the electrodes and a failure of the electroporation. The high conductance of DNA solutions prepared using prior art methods practically limits electroporation reactions to low DNA concentrations, since higher concentrations results in electrical arcing. The invention solves this problem, in part, by providing a method of electroporating cells with affinity purified DNA and/or at DNA concentrations much greater than was thought possible. It has been discovered that the DNA in prior art DNA preparations contributes only a small proportion of the total conductance; the majority of the conductance in these preparations is due to ionic impurities. The present invention uses affinity DNA purification to reduce ionic impurities and thus reduce the conductance associated with a unit mass of DNA. Although the prior art generally suggests using purified DNA for electroporation and several standard purifications have been used, for example, DNA precipitation and membrane filtration, the use of affinity purification has not been utilized and the very high DNA concentrations which can be used in the method of the invention and the resulting high transformation yields are surprising.

The invention provides an improved method of transforming cells by highly purifying DNA, for example recombinant clonable DNA, preferably closed circular DNA, more preferably phage or phagemid vector DNA. The invention enables one to prepare DNA solutions of high concentration, preferably an aqueous solution having very low conductance, for example a non-

buffer aqueous or water/glycerol solution at concentrations up to hundreds of micrograms of DNA per mL through the use of affinity purification of DNA to remove impurities which increase the conductance and shorten the time constant during electroporation. Electroporation using the higher DNA concentrations of this invention improves the transformation yield, but does not result in unacceptably higher cell death or loss of host cell viability. The method of the invention increases the amount of heterologous DNA, for example recombinant clonable DNA, which can be transformed into a cell. This increase in DNA entering the host cell provides a greater number of transformants per electroporation and allows one to prepare larger combinatorial libraries which overcomes the prior art problem of small library size using recombinant DNA.

The method of the invention also provides improved transformation yield using host cell concentrations higher than those used in the prior art to further improve transformation yield and combinatorial library size.

The invention also provides a novel *E. coli* strain containing a phage F′ factor which is particularly useful for the preparation of phage and phagemid libraries of variant peptides, proteins and antibodies for use in phage display systems.

BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A, 1B, and 1C: Protein VIII variants selected for increased display of fusion proteins. Figure 1A show a Zone-1 library encompassing protein VIII residues 1 through 10. Figure 1B shows a Zone-2 library encompassing protein VIII residues 11 through 20. Figure 1Cshows a Zone-3 library encompassing protein VIII residues 21 through 30. The possible variations at each position within the library are shown followed by wild-type and selected sequences. The DNA sequence is shown above in italics with the deduced amino acid sequence below in normal text (the amber stop codon (TAG) is suppressed as glutamine in *E. coli* XL1-Blue). DNA degeneracies are represented in the IUB code (K = G/T, N = A/C/G/T, R = A/G, S = G/C, W= A/T, Y = C/T). *Selected for streptavidin (SAV) display. All others were selected for hGH display.

Figure 2: Phage ELISAs for hGH display with protein VIII and selected protein VIII variants. The hGHbp was used as target (Kd = 1.6 nM, Pearce, K.H. Jr., et al. (1997) J. Biol. Chem 272:20595-20602). Display was measured for hGH-protein VIII (circles), hGH- protein VIII (1a) (squares), hGH- protein VIII (2a) (diamonds), and hGH- protein VIII (3a) (triangles). Phage were produced from cultures which were either uninduced (unfilled) or induced with 10 uM IPTG (filled). The sequences for the protein VIII variants are shown in Figure 1.

Figure 3: Phage ELISAs for display of hGH mutants. The hGHbp was used as target. Display was measured for wild-type hGH (circles, K_d = 1.6 nM), hGH(R64A) (squares, K_d = 13.8 nM), hGH(Y164A/R178A) (diamonds, K_d = 169 nM), and hGH(K172A/R178A), (triangles, K_d = 820 nM). hGH was fused to either wild-type protein VIII (unfilled) or protein VIII(1a) (filled). Phage were produced from cultures induced with 10 μM IPTG.

Figures 4A and 4B: Phage ELISAs for SAV display. Figure 4A shows results using anti-SAV polyclonal antibody as target. Figure 4B shows results using Biotin-BSA conjugate as target. Display was measured for SAV-protein VIII (circles), SAV-protein VIII (2e) (triangles), SAV-protein VIII (2f) (diamonds), and SAV- protein VIII (2a) (squares). Phage were produced from uninduced cultures.

Figure 5: Phage ELISAs for hGH display with protein VIII variants combining mutations in different zones. An anti-hGH monoclonal antibody was used as target. Display was measured for hGH displayed as a fusion with protein VIII (filled circles), protein VIII(1a) (filled squares), protein VIII(2a) (filled diamonds), protein VIII(3a) (filled triangles), protein VIII containing the mutations of protein VIII(1a) and protein VIII(2a) (unfilled circles), or protein VIII containing the mutations of protein VIII(2a) and protein VIII(3a) (unfilled squares).

Figure 6: Phage ELISAs for hGH display with protein VIII variants derived from protein VIII(2a). An anti-hGH monoclonal antibody was used as target. Display was measured for hGH displayed as a fusion with either protein VIII(2a) (filled circles) or with protein VIII(2a) containing the following mutation: E12N (unfilled squares), D16A (unfilled circles), or I17S (unfilled triangles).

Figures 7A and 7B: Linkers selected for hGH display or SAV display. 7A) Linkers selected for display of hGH. Linkers were of the form $(Gly)_3(Xaa)_{14}(Gly)_2$, where (Xaa)14 is the selected sequence shown. 7B) Linkers selected for display of SAV.

Figures 8A and 8B: Phage ELISAs for protein display with selected linkers. 8A) hGH displayed on protein VIII using either a Gly/Ser linker (phagemid pS349, circles) or the linker selectant Link1 (Fig. 7A, squares). Phage were produced from cultures which were either uninduced (unfilled) or induced with 10 uM IPTG (filled). The hGHbp was used as target. 8B) SAV displayed on wild type protein VIII using a Gly/Ser linker (unfilled circles) or on variant protein VIII(2e) using either a Gly/Ser linker (unfilled squares), link18 (filled circles), link29 (filled squares), link34 (filled diamonds), or link37 (filled triangles). The Gly/Ser sequence was identical to the Gly/Ser linker encoded by pS349. The sequences of the other linkers are shown in Figure 7B. Biotinylated BSA was used as target.

Figure 9: Phage ELISAs for hGH display with protein VIII and selected protein VIII variants. The hGHbp was used as target (Kd = 1.6 nM, Pearce, K.H. Jr. et al., (1997) J. Biol. Chem. 272:20595-20602). Display was measured for hGH-protein VIII expressed from phagemid pS1607 (see Example 11) (circles), hGH-protein VIII(1a) (squares), hGH-protein VIII(2a) (diamonds), and hGH-protein VIII(3a) (triangles). Phage were produced from cultures which were either uninduced (unfilled) or induced with 10 uM IPTG (filled). The sequences for the protein VIII variants are show in Figure 1.

Figure 10: Phage ELISAs for Fab display with protein VIII and a protein VIII mutant. A monoclonal antibody specific for a peptide flag fused to the N-terminus of the Fab heavy chain was

used as target. Display was measured for Fab-protein VIII (circles) or Fab-protein VIII(S13A/S17I).

Figure 11: Phage ELISAs for hGH display with protein VIII variants combining mutations in different zones. An anti-hGH monoclonal antibody was used as target. hGH was fused to wild-type protein VIII (wt), a protein VIII selectant from zone 1, 2, or 3 (1a, 2a, or 3a, respectively) or protein VIII variants combining mutations from these selectants (*e.g.* 1a+2a combines 1a residues 1 through 10 with 2a residues 11 through 20 and wild type residues 21 through 50). Sequences for the protein VIII selectants are shown in Figure 1.

Figure 12: Site directed mutagenesis of protein VIII variants highlights key positions for enhanced display and enables modulated hGH display. hGH was fused to variants derived from A) zone 1 selectant 1a, B) zone 2 selectant 2a, or C) zone 3 selectant 3a. For each protein VIII selectant, the effect of every possible single back mutation to the wild-type sequence is shown (*e.g.* D1A indicates the protein VIII variant obtained by introducing the mutation D1A into selectant 1a. In addition, D) double and triple back mutations were introduced into zone 2 selectant 2a to further modulate hGH display. Sequences for the protein VIII selectants are shown in Figure 1.

Figure 13: Phage ELISAs for the display of a peptide fused to the C-terminus of protein VIII using poly-glycine linkers. A hexapeptide (HHHHHA, referred to as a pentaHis flag) was fused to the C-terminus of protein VIII with intervening linkers containing varying numbers of Gly residues as indicated (linker length, X-axis). There is a large increase in display when the linker length is increased from eight to nine residues. The phage were used at a concentration of 2×10^{12} phage/mL. An anti-(His)5 antibody (Qiagen) was used as the capture target. See Example 22.

Figure 14: Phage ELISAs for the display of a peptide fused to the C-terminus of protein VIII using an optimized linker sequence. A polyHis flag was fused to the C-terminus of protein VIII with intervening linkers as follows: (Gly)8 (open circles), (Gly)9 (squares), (Gly)10 (diamonds), (Gly)12 (triangles), or optimized linker-1 (filled circles). The highest levels of display were observed with the optimized linker. Phage were serially diluted 5-fold from a starting concentration of 10^{13} phage/mL. An anti-(His)5 antibody (Qiagen) was used as the capture target. See Examples 22 and 23.

Figure 15: P12 variants selected for the display of a polyHis flag as a C-terminal fusion. The variable region of each P12 is shown. The complete sequence for each P12 is as follows: MSKSTFKKFLK-(x)19-ETASAQLSNFAAKAPDDGEA (SEQ ID NO. 1).
Where "(x)19" is the nineteen residue sequence inserted in the library construction as shown in the figure. The possible variations at each position within the library are shown followed by the selected sequences. The DNA sequence is shown with the deduced amino acid sequence below. The numerical designation for each sequence is shown to the left. The numbering above refers to the position of each codon within the nineteen residue library insertion. See Example 24.

Figure 16: Phage ELISAs for the display of a polyHis flag as a C-terminal fusion with P12 variants. An anti-(His)4 antibody was used as the capture target (unfilled bars). As a negative

control, phage binding to a BSA-blocked plate was also measured (filled bars). The phage were used at a concentration of 10^{13} phage/mL. See Example 24.

Figure 17: Phage ELISAs for hGH or hGHsm display with P12-7. Display was measured for hGHsm fused to the C-terminus of P12-7 (phagemid pS1258, open circles), hGH fused to the C-terminus of P12-7 (phagemid pW930a, open squares), hGHsm fused to the C-terminus of P12-1 (phagemid pS1239b, filled circles), hGH fused to the C-terminus of P12-1 (phagemid pS1239a, filled squares), or hGH fused to the N-terminus of protein VIII (phagemid pS1607, filled diamonds). An anti-hGH monoclonal antibody was used as target. See Example 25.

Figure 18: Phage ELISAs for the display of a peptide fused to the C-terminus of the protein III C-terminal domain using selected linkers. A hexaHis flag was displayed with intervening linker sequences as follows: linker-g3-1 (open circles), linker-g3-2 (open squares), linker-g3-3, (open diamonds). Display was also measured for polyHis flags displayed as either N-terminal fusions with protein VIII (filled circles) or as C-terminal fusions with protein VIII using optimized linker-1 (filled diamonds). Phage derived from a phagemid not encoding a polyHis flag were also included as a negative control (filled squares). See Example 26.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

DEFINITIONS

The term "affinity purification" means the purification of a molecule based on a specific attraction or binding of the molecule to a chemical or binding partner to form a combination or complex which allows the molecule to be separated from impurities while remaining bound or attracted to the partner moiety.

The term "antibody" is used in the broadest sense and specifically covers single monoclonal antibodies (including agonist and antagonist antibodies), antibody compositions with polyepitopic specificity, affinity matured antibodies, humanized antibodies, chimeric antibodies, as well as antibody fragments (*e.g.*, Fab, $F(ab')_2$, scFv and Fv), so long as they exhibit the desired biological activity. An affinity matured antibody will typically have its binding affinity increased above that of the isolated or natural antibody or fragment thereof by from 2 to 500 fold. Preferred affinity matured antibodies will have nanomolar or even picomolar affinities to the receptor antigen. Affinity matured antibodies are produced by procedures known in the art. Marks, J. D. *et al. Bio/Technology* 10:779-783 (1992) describes affinity maturation by VH and VL domain shuffling. Random mutagenesis of CDR and/or framework residues is described by: Barbas, C. F. *et al. Proc Nat. Acad. Sci, USA* 91:3809-3813 (1994), Schier, R. *et al. Gene* 169:147-155 (1995), Yelton, D. E. *et al. J. Immunol.* 155:1994-2004 (1995), Jackson, J.R. *et al., J. Immunol.* 154(7):3310-9 (1995), and Hawkins, R.E. *et al, J. Mol. Biol.* 226:889-896 (1992). Humanized antibodies are known. Jones *et al., Nature*, 321:522-525 (1986); Reichmann *et al., Nature*, 332:323-329 (1988); and Presta, *Curr. Op. Struct. Biol.*, 2:593-596 (1992)).

An "Fv" fragment is the minimum antibody fragment which contains a complete antigen recognition and binding site. This region consists of a dimer of one heavy and one light chain variable domain in tight, non-covalent association. It is in this configuration that the three CDRs of each variable domain interact to define an antigen binding site on the surface of the V_H-V_L dimer. Collectively, the six CDRs confer antigen binding specificity to the antibody. However, even a single variable domain (or half of an Fv comprising only three CDRs specific for an antigen) has the ability to recognize and bind antigen, although at a lower affinity than the entire binding site.

The "Fab" fragment also contains the constant domain of the light chain and the first constant domain (CH1) of the heavy chain. Fab' fragments differ from Fab fragments by the addition of a few residues at the carboxy terminus of the heavy chain CH1 domain including one or more cysteines from the antibody hinge region. Fab'-SH is the designation herein for Fab' in which the cysteine residue(s) of the constant domains bear a free thiol group. $F(ab')_2$ antibody fragments originally were produced as pairs of Fab' fragments which have hinge cysteines between them. Other, chemical couplings of antibody fragments are also known.

"Single-chain Fv" or "sFv" antibody fragments comprise the V_H and V_L domains of antibody, wherein these domains are present in a single polypeptide chain. Generally, the Fv polypeptide further comprises a polypeptide linker between the V_H and V_L domains which enables the sFv to form the desired structure for antigen binding. For a review of sFv see Pluckthun in *The Pharmacology of Monoclonal Antibodies*, vol. 113, Rosenburg and Moore eds. Springer-Verlag, New York, pp. 269-315 (1994).

The term "diabodies" refers to small antibody fragments with two antigen-binding sites, which fragments comprise a heavy chain variable domain (V_H) connected to a light chain variable domain (V_L) in the same polypeptide chain (V_H - V_L). By using a linker that is too short to allow pairing between the two domains on the same chain, the domains are forced to pair with the complementary domains of another chain and create two antigen-binding sites. Diabodies are described more fully in, for example, EP 404,097; WO 93/11161; and Hollinger *et al., Proc. Natl. Acad. Sci. USA* 90:6444-6448 (1993).

The expression "linear antibodies" refers to the antibodies described in Zapata *et al. Protein Eng.* 8(10):1057-1062 (1995). Briefly, these antibodies comprise a pair of tandem Fd segments (V_H-C_H1-V_H-C_H1) which form a pair of antigen binding regions. Linear antibodies can be bispecific or monospecific.

"Cell," "cell line," and "cell culture" are used interchangeably herein and such designations include all progeny of a cell or cell line. Thus, for example, terms like "transformants" and "transformed cells" include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant

progeny that have the same function or biological activity as screened for in the originally transformed cell are included. Where distinct designations are intended, it will be clear from the context.

The terms "competent cells" and "electoporation competent cells" mean cells which are in a state of competence and able to take up DNAs from a variety of sources. The state may be transient or permanent. Electroporation competent cells are able to take up DNA during electroporation.

"Control sequences" when referring to expression means DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, a ribosome binding site, and possibly, other as yet poorly understood sequences. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

The term "coat protein" means a protein, at least a portion of which is present on the surface of the virus particle. From a functional perspective, a coat protein is any protein which associates with a virus particle during the viral assembly process in a host cell, and remains associated with the assembled virus until it infects another host cell. The coat protein may be the major coat protein or may be a minor coat protein. A "major" coat protein is a coat protein which is present in the viral coat at 10 copies of the protein or more. A major coat protein may be present in tens, hundreds or even thousands of copies per virion.

The "detection limit" for a chemical entity in a particular assay is the minimum concentration of that entity which can be detected above the background level for that assay. For example, in the phage ELISA of Example 5, the "detection limit" for a particular phage displaying a particular protein (e.g. hGH) is the phage concentration at which the particular phage produces an ELISA signal above that produced by a control phage not displaying the protein.

The terms "electroporation" and "electroporating" mean a process in which foreign matter (protein, nucleic acid, etc.) is introduced into a cell by applying a voltage to the cell under conditions sufficient to allow uptake of the foreign matter into the cell. The foreign matter is typically DNA.

An "F factor" or "F′ episome" is a DNA which, when present in a cell, allows bacteriophage to infect the cell. The episome may contain other genes, for example selection genes, marker genes, etc. Common F′ episomes are found in well known *E. coli* strains including CJ236, CSH18, DH5alphaF′, JM101 (same as in JM103, JM105, JM107, JM109, JM110), KS1000, XL1-BLUE and 71-18. These strains and the episomes contained therein are commercially available (New England Biolabs) and many have been deposited in recognized depositories such as ATCC in Manassas, VA.

A "fusion protein" is a polypeptide having two portions covalently linked together, where each of the portions is a polypeptide having a different property. The property may be a biological

property, such as activity in vitro or in vivo. The property may also be a simple chemical or physical property, such as binding to a target molecule, catalysis of a reaction, etc. The two portions may be linked directly by a single peptide bond or through a peptide linker containing one or more amino acid residues. Generally, the two portions and the linker will be in reading frame with each other.

"Heterologous DNA" is any DNA that is introduced into a host cell. The DNA may be derived from a variety of sources including genomic DNA, cDNA, synthetic DNA and fusions or combinations of these. The DNA may include DNA from the same cell or cell type as the host or recipient cell or DNA from a different cell type, for example, from a mammal or plant. The DNA may, optionally, include selection genes, for example, antibiotic resistance genes, temperature resistance genes, etc.

"Ligation" is the process of forming phosphodiester bonds between two nucleic acid fragments. For ligation of the two fragments, the ends of the fragments must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary first to convert the staggered ends commonly produced after endonuclease digestion to blunt ends to make them compatible for ligation. For blunting the ends, the DNA is treated in a suitable buffer for at least 15 minutes at 15°C with about 10 units of the Klenow fragment of DNA polymerase I or T4 DNA polymerase in the presence of the four deoxyribonucleotide triphosphates. The DNA is then purified by phenol-chloroform extraction and ethanol precipitation. The DNA fragments that are to be ligated together are put in solution in about equimolar amounts. The solution will also contain ATP, ligase buffer, and a ligase such as T4 DNA ligase at about 10 units per 0.5 μg of DNA. If the DNA is to be ligated into a vector, the vector is first linearized by digestion with the appropriate restriction endonuclease(s). The linearized fragment is then treated with bacterial alkaline phosphatase or calf intestinal phosphatase to prevent self-ligation during the ligation step.

A "mutation" is a deletion, insertion, or substitution of a nucleotide(s) relative to a reference nucleotide sequence, such as a wild type sequence.

A "silent mutation" is a mutation which does not change the amino acid sequence of the translated polypeptide product of a given DNA sequence.

A "non-silent mutation" is a mutation which changes the amino acide sequence of the translated polypeptide product of a given DNA sequence.

"Operably linked" when referring to nucleic acids means that the nucleic acids are placed in a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation.

Generally, "operably linked" means that the DNA sequences being linked are contiguous and, in the case of a secretory leader, contiguous and in reading phase. However, enhancers do not have to be contiguous. Linking is accomplished by ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adapters or linkers are used in accord with conventional practice.

"Phage display" is a technique by which variant polypeptides are displayed as fusion proteins to a coat protein on the surface of phage, e.g. filamentous phage, particles. A utility of phage display lies in the fact that large libraries of randomized protein variants can be rapidly and efficiently sorted for those sequences that bind to a target molecule with high affinity. Display of peptides and proteins libraries on phage has been used for screening millions of polypeptides for ones with specific binding properties. Polyvalent phage display methods have been used for displaying small random peptides and small proteins through fusions to either gene III or gene VIII of filamentous phage. Wells and Lowman, *Curr. Opin. Struct. Biol.*, 1992, 3:355-362 and references cited therein. In monovalent phage display, a protein or peptide library is fused to a gene III or a portion thereof and expressed at low levels in the presence of wild type gene III protein so that phage particles display one copy or none of the fusion proteins. Avidity effects are reduced relative to polyvalent phage so that sorting is on the basis of intrinsic ligand affinity, and phagemid vectors are used, which simplify DNA manipulations. Lowman and Wells, *Methods: A companion to Methods in Enzymology*, 1991, 3:205-216.

A "phagemid" is a plasmid vector having a bacterial origin of replication, e.g., ColE1, and a copy of an intergenic region of a bacteriophage. The phagemid may be based on any known bacteriophage, including filamentous bacteriophage and lambdoid bacteriophage. The plasmid will also generally contain a selectable marker for antibiotic resistance. Segments of DNA cloned into these vectors can be propagated as plasmids. When cells harboring these vectors are provided with all genes necessary for the production of phage particles, the mode of replication of the plasmid changes to rolling circle replication to generate copies of one strand of the plasmid DNA and package phage particles. The phagemid may form infectious or non-infectious phage particles. This term includes phagemids which contain a phage coat protein gene or fragment thereof linked to a heterologous polypeptide gene as a gene fusion such that the heterologous polypeptide is displayed on the surface of the phage particle. Sambrook et al., above, 4.17.

The term "phage vector" means a double stranded replicative form of a bacteriophage containing a heterologous gene and capable of replication. The phage vector has a phage origin of replication allowing phage replication and phage particle formation. The phage is preferably a filamentous bacteriophage, such as an M13, f1, fd, Pf3 phage or a derivative thereof, or a lambdoid phage, such as lambda, 21, phi80, phi81, 82, 424, 434, etc., or a derivative thereof.

"Preparation" of DNA from cells means isolating the plasmid DNA from a culture of the host cells. Commonly used methods for DNA preparation are the large- and small-scale plasmid preparations described in sections 1.25-1.33 of Sambrook *et al.*, *supra*. After preparation of the

DNA, it can be purified by methods well known in the art such as that described in section 1.40 of Sambrook *et al.*, *supra*.

"Oligonucleotides" are short-length, single- or double-stranded polydeoxynucleotides that are chemically synthesized by known methods (such as phosphotriester, phosphite, or phosphoramidite chemistry, using solid-phase techniques such as described in EP 266,032 published 4 May 1988, or via deoxynucleoside H-phosphonate intermediates as described by Froehler *et al.*, *Nucl. Acids Res.*, 14:5399-5407 (1986)). Further methods include the polymerase chain reaction defined below and other autoprimer methods and oligonucleotide syntheses on solid supports. All of these methods are described in Engels *et al.*, *Agnew. Chem. Int. Ed. Engl.*, 28:716-734 (1989). These methods are used if the entire nucleic acid sequence of the gene is known, or the sequence of the nucleic acid complementary to the coding strand is available. Alternatively, if the target amino acid sequence is known, one may infer potential nucleic acid sequences using known and preferred coding residues for each amino acid residue. The oligonucleotides are then purified on polyacrylamide gels.

"Polymerase chain reaction" or "PCR" refers to a procedure or technique in which minute amounts of a specific piece of nucleic acid, RNA and/or DNA, are amplified as described in U.S. Patent No. 4,683,195 issued 28 July 1987. Generally, sequence information from the ends of the region of interest or beyond needs to be available, such that oligonucleotide primers can be designed; these primers will be identical or similar in sequence to opposite strands of the template to be amplified. The 5′ terminal nucleotides of the two primers may coincide with the ends of the amplified material. PCR can be used to amplify specific RNA sequences, specific DNA sequences from total genomic DNA, and cDNA transcribed from total cellular RNA, bacteriophage or plasmid sequences, *etc.* See generally Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.*, 51:263 (1987); Erlich, ed., *PCR Technology*, (Stockton Press, NY, 1989). As used herein, PCR is considered to be one, but not the only, example of a nucleic acid polymerase reaction method for amplifying a nucleic acid test sample comprising the use of a known nucleic acid as a primer and a nucleic acid polymerase to amplify or generate a specific piece of nucleic acid.

DNA is "purified" when the DNA is separated from non-nucleic acid impurities. The impurities may be polar, non-polar, ionic, etc.

"Recovery" or "isolation" of a given fragment of DNA from a restriction digest means separation of the digest on polyacrylamide or agarose gel by electrophoresis, identification of the fragment of interest by comparison of its mobility versus that of marker DNA fragments of known molecular weight, removal of the gel section containing the desired fragment, and separation of the gel from DNA. This procedure is known generally. For example, see Lawn *et al.*, *Nucleic Acids Res.*, 9:6103-6114 (1981), and Goeddel *et al.*, *Nucleic Acids Res.*, 8:4057 (1980).

A chemical group or species having a "specific binding affinity for DNA" means a molecule or portion thereof which forms a non-covalent bond with DNA which is stronger than the bonds formed with other celular components including proteins, salts, and lipids.

A "survivor" is a cell which remains viable after a transformation process.

A "transcription regulatory element" will contain one or more of the following components: an enhancer element, a promoter, an operator sequence, a repressor gene, and a transcription termination sequence. These components are well known in the art. U.S. 5,667,780.

A "transformant" is a cell which has taken up and maintained DNA as evidenced by the expression of a phenotype associated with the DNA (e.g., antibiotic resistance conferred by a protein encoded by the DNA).

"Transformation" means a process whereby a cell takes up DNA and becomes a "transformant". The DNA uptake may be permanent or transient.

"Transformation efficiency" means the number of transformants produced per unit mass of DNA following a transformation procedure (e.g. transformants per microgram of DNA).

"Transformation frequency" means the ratio of the number of transformants to the number of survivors.

"Transformation yield" means the number of transformants produced in a single electroporation reaction.

A "variant" or "mutant" of a starting polypeptide, such as a fusion protein or a heterologous polypeptide (heterologous to a phage), is a polypeptide that 1) has an amino acid sequence different from that of the starting polypeptide and 2) was derived from the starting polypeptide through either natural or artificial (manmade) mutagenesis. Such variants include, for example, deletions from, and/or insertions into and/or substitutions of, residues within the amino acid sequence of the polypeptide of interest. Any combination of deletion, insertion, and substitution may be made to arrive at the final variant or mutant construct, provided that the final construct possesses the desired functional characteristics. The amino acid changes also may alter post-translational processes of the polypeptide, such as changing the number or position of glycosylation sites. Methods for generating amino acid sequence variants of polypeptides are described in U. S. 5,534,615, expressly incorporated herein by reference.

Generally, a variant coat protein will possess at least 20% or 40% sequence identity and up to 70% or 85% sequence identity, more preferably up to 95% or 99.9% sequence identity, with the wild type coat protein. Percentage sequence identity is determined, for example, by the Fitch *et al.*, *Proc. Natl. Acad. Sci. USA* 80:1382-1386 (1983), version of the algorithm described by Needleman *et al.*, *J. Mol. Biol.* 48:443-453 (1970), after aligning the sequences to provide for maximum homology. Amino acid sequence variants of a polypeptide are prepared by introducing appropriate nucleotide changes into DNA encoding the polypeptide, or by peptide synthesis. An "altered residue" is a deletion, insertion or substitution of an amino acid residue relative to a reference amino acid sequence, such as a wild type sequence.

A "functional" mutant or variant is one which exhibits a detectable activity or function which is also detectably exhibited by the wild type protein. For example, a "functional" mutant or variant of the major coat protein is one which is stably incorporated into the phage coat at levels

which can be experimentally detected. Preferably, the phage coat incorporation can be detected in a range of about 1 fusion per 1000 virus particles up to about 1000 fusions per virus particle.

A "hyper-functional" mutant or variant is a functional mutant or variant whose activity exceeds that of the wild type. For example, a hyper-functional mutant or variant of the major coat protein is one which is stably incorporated into the phage coat at levels greater than those of the wild type protein in an identical context.

A "hypo-functional" mutant or variant is a functional mutant or variant whose activity is less than that of the wild type. For example, a hypo-functional mutant or variant of the major coat protein is one which is stably incorporated into the phage coat at levels less than those of the wild type protein in an identical context.

A "wild type" sequence or the sequence of a "wild type" protein, such as a coat protein, is the reference sequence from which variant polypeptides are derived through the introduction of mutations. In general, the "wild type" sequence for a given protein is the sequence that is most common in nature. Similarly, a "wild type" gene sequence is the sequence for that gene which is most commonly found in nature. Mutations may be introduced into a "wild type" gene (and thus the protein it encodes) either through natural processes or through man induced means. The products of such processes are "variant" or "mutant" forms of the original "wild type" protein or gene.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

A. Novel methods and cells

The present invention provides a method of transforming cells by electroporating competent cells in the presence of heterologous DNA, where the DNA has been purified by DNA affinity purification. Preferably, for library construction in bacteria, the DNA is present at a concentration of 25 micrograms/mL or greater. Preferably, the DNA is present at a concentration of about 30 micrograms/mL or greater, more preferably at a concentration of about 70 micrograms/mL or greater and even more preferably at a concentration of about 100 micrograms/mL or greater even up to several hundreds of micrograms/mL. Generally, the method of the invention will utilize DNA concentrations in the range of about 50 to about 500 micrograms/mL. It has been discovered that by highly purifying the heterologous DNA, a time constant during electroporation greater than 3.0 milliseconds (ms) is possible even when the DNA concentration is very high, which results in a high transformation efficiency. Over the DNA concentration range of about 50 microgram/mL to about 400 microgram/mL, the method of the invention allows the use of time constants in the range of about 3.6 to about 4.4 ms using standard electroporation instruments. The invention therefore provides a method with greater dynamic range in DNA concentration than previously known.

The high DNA concentrations used in the method of the invention are obtained by highly purifying DNA used to transform the competent cells. In the method of the invention, the DNA is purified to remove contaminants which increase the conductance of the DNA solution used in the electroporating process. The DNA may be purified by any known method, however, a preferred purification method is the use of DNA affinity purification. The purification of DNA, e.g., recombinant linear or plasmid DNA, using DNA binding resins and affinity reagents is well known and any of the known methods can be used in this invention (Vogelstein, B. and Gillespie, D., 1979, *Proc. Natl. Acad. Sci. USA*, 76:615; Callen, W., 1993, Strategies, 6:52-53). Commercially available DNA isolation and purification kits are also available from several sources including Stratagene (CLEARCUT Miniprep Kit), and Life Technologies (GLASSMAX DNA Isolation Systems). Suitable nonlimiting methods of DNA purification include column chromatography (U.S. 5,707,812), the use of hydroxylated silical polymers (U.S. 5,693,785), rehydrated silica gel (U.S. 4,923,978), boronated silicates (U.S. 5,674,997), modified glass fiber membranes (U.S. 5,650,506; U.S. 5,438,127), fluorinated adsorbents (U.S. 5,625,054; U.S. 5,438,129), diatomaceous earth (U.S. 5,075,430), dialysis (U.S. 4,921,952), gel polymers (U.S. 5,106,966) and the use of chaotropic compounds with DNA binding reagents (U.S. 5,234,809). After purification, the DNA is eluted or otherwise resuspended in water, preferably distilled or deionized water, for use in electroporation at the concentrations of the invention. The use of low salt buffer solutions is also contemplated where the solution has low electrical conductivity, i.e., is compatible with the use of the high DNA concentrations of the invention with time constants greater than about 3.0 ms.

Any cells which can be transformed by electroporation may be used as host cells in the method of the present invention. Suitable host cells which can be transformed with heterologous DNA in the method of the invention include animal cells (Neumann et al., *EMBO J.*, (1982), 1:841; Wong and Neumann, *Biochem. Biophys. Res. Commun.*, (1982), 107:584; Potter et al., *Proc. Natl. Acad. Sci., USA*, (1984) 81:7161; Sugden et al., *Mol. Cell. Biol.*, (1985), 5:410; Toneguzzo et al., *Mol. Cell. Biol.*, (1986), 6:703; Pur-Kaspa et al., *Mol. Cell. Biol.*, (1986), 6:716), plant cells (Fromm et al., *Proc. Natl. Acad. Sci., USA*, (1985), 82:5824; Fromm et al., *Nature*, (1986), 319:791; Ecker and Davis, *Proc. Natl. Acad. Sci., USA*, (1986) 83:5372) and bacterial cells (Chu et al., *Nucleic Acids Res.*, (1987), 15:1311; Knutson and Yee, *Anal. Biochem.*, (1987), 164:44). Prokaryotes are the preferred host cells for this invention. See also Andreason and Evans, *Biotechniques*, (1988), 6:650 which describes parameters which effect transfection efficiencies for varying cell lines. Suitable bacterial cells include *E. coli* (Dower et al., above; Taketo, *Biochim. Biophys. Acta*, (1988), 149:318), *L. casei* (Chassy and Flickinger, *FEMS Microbiol. Lett.*, (1987), 44:173), *Strept. lactis* (Powell et al., *Appl. Environ. Microbiol.*, (1988), 54:655; Harlander, *Streptococcal Genetics*, ed . J. Ferretti and R. Curtiss, III), page 229, American Society for Microbiology, Washington, D.C., (1987)), *Strept. thermophilus* (Somkuti and Steinberg, *Proc. 4th Eur. Cong. Biotechnology*, 1987, 1:412); *Campylobacter jejuni* (Miller et al., *Proc. Natl. Acad. Sci., USA*, (1988) 85:856), and other bacterial strains (Fielder and Wirth, *Anal. Biochem.*, (1988), 170:38) including bacilli such as *Bacillus subtilis*, other enterobacteriaceae such as *Salmonella typhimurium* or *Serratia marcesans*, and various *Pseudomonas* species which may all be used as hosts. Suitable *E. coli* strains include JM101, *E. coli* K12 strain 294 (ATCC number 31,446), *E. coli* strain W3110 (ATCC number 27,325), *E. coli* X1776 (ATCC number 31,537), *E. coli* XL-1Blue (Stratagene), and *E. coli* B; however many other strains of *E. coli*, such as XL1-Blue MRF', SURE, ABLE C, ABLE K, WM1100, MC1061, HB101, CJ136, MV1190, JS4, JS5, NM522, NM538, NM539, TG1and many other species and genera of prokaryotes may be used as well.

Cells are made competent using known procedures. Sambrook et al., above, 1.76-1.81, 16.30.

The heterologous DNA is preferably in the form of a replicable transcription or expression vector, such as a plasmid, phage or phagemid which can be constructed with relative ease and readily amplified. These vectors generally contain a promoter, a signal sequence, phenotypic selection genes, origins of replication, and other necessary components which are known to those of ordinary skill in this art. Construction of suitable vectors containing these components as well as the gene encoding one or more desired cloned polypeptides are prepared using standard recombinant DNA procedures as described in Sambrook et al., above. Isolated DNA fragments to be combined to form the vector are cleaved, tailored, and ligated together in a specific order and orientation to generate the desired vector.

The gene encoding the desired polypeptide (i.e., a peptide or a polypeptide with a rigid secondary structure) can be obtained by methods known in the art (see generally, Sambrook *et al.*). If the sequence of the gene is known, the DNA encoding the gene may be chemically synthesized (Merrfield, *J. Am. Chem. Soc.*, 85 :2149 (1963)). If the sequence of the gene is not known, or if the gene has not previously been isolated, it may be cloned from a cDNA library (made from RNA obtained from a suitable tissue in which the desired gene is expressed) or from a suitable genomic DNA library. The gene is then isolated using an appropriate probe. For cDNA libraries, suitable probes include monoclonal or polyclonal antibodies (provided that the cDNA library is an expression library), oligonucleotides, and complementary or homologous cDNAs or fragments thereof. The probes that may be used to isolate the gene of interest from genomic DNA libraries include cDNAs or fragments thereof that encode the same or a similar gene, homologous genomic DNAs or DNA fragments, and oligonucleotides. Screening the cDNA or genomic library with the selected probe is conducted using standard procedures as described in chapters 10-12 of Sambrook et al., above.

An alternative means to isolating the gene encoding the protein of interest is to use polymerase chain reaction methodology (PCR) as described in section 14 of Sambrook et al., above. This method requires the use of oligonucleotides that will hybridize to the gene of interest; thus, at least some of the DNA sequence for this gene must be known in order to generate the oligonucleotides.

After the gene has been isolated, it may be inserted into a suitable vector (preferably a plasmid) for amplification, as described generally in Sambrook et al.

The DNA is cleaved using the appropriate restriction enzyme or enzymes in a suitable buffer. In general, about 0.2-1 μg of plasmid or DNA fragments is used with about 1-2 units of the appropriate restriction enzyme in about 20 μl of buffer solution. Appropriate buffers, DNA concentrations, and incubation times and temperatures are specified by the manufacturers of the restriction enzymes. Generally, incubation times of about one or two hours at 37°C are adequate, although several enzymes require higher temperatures. After incubation, the enzymes and other contaminants are removed by extraction of the digestion solution with a mixture of phenol and chloroform, and the DNA is recovered from the aqueous fraction by precipitation with ethanol or other DNA purification technique.

To ligate the DNA fragments together to form a functional vector, the ends of the DNA fragments must be compatible with each other. In some cases, the ends will be directly compatible after endonuclease digestion. However, it may be necessary to first convert the sticky ends commonly produced by endonuclease digestion to blunt ends to make them compatible for ligation. To blunt the ends, the DNA is treated in a suitable buffer for at least 15 minutes at 15 C with 10 units of the Klenow fragment of DNA polymerase I (Klenow) in the presence of the four deoxynucleotide triphosphates. The DNA is then purified by phenol-chloroform extraction and ethanol precipitation or other DNA purification technique.

The cleaved DNA fragments may be size-separated and selected using DNA gel electrophoresis. The DNA may be electrophoresed through either an agarose or a polyacrylamide matrix. The selection of the matrix will depend on the size of the DNA fragments to be separated. After electrophoresis, the DNA is extracted from the matrix by electroelution, or, if low-melting agarose has been used as the matrix, by melting the agarose and extracting the DNA from it, as described in sections 6.30-6.33 of Sambrook et al., *supra*.

The DNA fragments that are to be ligated together (previously digested with the appropriate restriction enzymes such that the ends of each fragment to be ligated are compatible) are put in solution in about equimolar amounts. The solution will also contain ATP, ligase buffer and a ligase such as T4 DNA ligase at about 10 units per 0.5 μg of DNA. If the DNA fragment is to be ligated into a vector, the vector is at first linearized by cutting with the appropriate restriction endonuclease(s). The linearized vector is then treated with alkaline phosphatase or calf intestinal phosphatase. The phosphatasing prevents self-ligation of the vector during the ligation step.

After ligation, the vector with the foreign gene now inserted is purified as described above and transformed into a suitable host cell such as those described above by electroporation using known and commercially available electroporation instruments and the procedures outlined by the manufacturers and described generally in Dower et al., above. The invention provides high transformation yields, a single electroporation reaction typically yields greater than 1×10^{10} transformants. However, more than one (a plurality) electroporation may be conducted to increase the amount of DNA which is transformed into the host cells. Repeated electroporations are conducted as described in the art. See Vaughan et al., above. The number of additional electroporations may vary as desired from several (2,3,4,...10) up to tens (10, 20, 30,...100) and even hundreds (100, 200, 300,...1000). Repeated electroporations may be desired to increase the size of a combinatorial library, e.g. an antibody library, transformed into the host cells. With a plurality of electroporations, it is possible to produce a library having at least 1.0×10^{12}, even 2.0×10^{12}, different members (clones, DNA vectors such as phage, phagemids, plasmids, etc., cells, etc.).

Electroporation may be carried out using methods known in the art and described, for example, in U.S. 4,910,140; U.S. 5,186,800; U.S. 4,849,355; , U.S. 5,173,158; U.S. 5,098,843; U.S. 5,422,272; U.S. 5,232,856; U.S. 5,283,194; U.S. 5,128,257; U.S. 5,750,373; U.S. 4,956,288 or any other known batch or continuous electroporation process together with the improvements of the invention.

Typically, electrocompetent cells are mixed with a solution of DNA at the desired concentration at ice temperatures. An aliquot of the mixture is placed into a cuvette and placed in an electroporation instrument, e.g., GENE PULSER (Biorad) having a typical gap of 0.2 cm. Each cuvette is electroporated as described by the manufacturer. Typical settings are: voltage = 2.5 kV, resistance = 200 ohms, capacitance = 25 mF. The cuvette is then immediately removed, SOC media (Maniatis) is added, and the sample is transferred to a 250 mL baffled flask. The contents of

several cuvettes may be combined after electroporation. The culture is then shaken at 37°C to culture the transformed cells.

The transformed cells are generally selected by growth on an antibiotic, commonly tetracycline (tet) or ampicillin (amp), to which they are rendered resistant due to the presence of tet and/or amp resistance genes in the vector.

After selection of the transformed cells, these cells are grown in culture and the vector DNA (plasmid or other vector with the foreign gene inserted) may then be isolated. Vector DNA can be isolated using methods known in the art. Two suitable methods are the small scale preparation of DNA and the large-scale preparation of DNA as described in sections 1.25-1.33 of Sambrook et al., *supra*. The isolated DNA can be purified by methods known in the art such as that described in section 1.40 of Sambrook et al., above and as described above.. This purified DNA is then analyzed by restriction mapping and/or DNA sequencing. DNA sequencing is generally performed by either the method of Messing et al., *Nucleic Acids Res.*, 9:309 (1981) or by the method of Maxam et al., *Meth. Enzymol.*, 65:499 (1980).

This invention also contemplates fusing the gene encoding the desired polypeptide (gene 1) to a second gene (gene 2) such that a fusion protein is generated during transcription. Gene 2 is typically a coat protein gene of a filamentous phage, preferably phage M13 or a related phage, and the gene is preferably the coat protein III gene or the coat protein VIII gene, or a fragment thereof. See U.S. 5,750,373; WO 95/34683. Fusion of genes 1 and 2 may be accomplished by inserting gene 2 into a particular site on a plasmid that contains gene 1, or by inserting gene 1 into a particular site on a plasmid that contains gene 2 using the standard techniques described above.

Alternatively, gene 2 may be a molecular tag for identifying and/or capturing and purifying the transcribed fusion protein. For example, gene 2 may encode for Herpes simplex virus glycoprotein D (Paborsky et al., 1990, *Protein Engineering*, 3:547-553) which can be used to affinity purify the fusion protein through binding to an anti-gD antibody. Gene 2 may also code for a polyhistidine, e.g., $(his)_6$ (Sporeno et al., 1994, *J. Biol. Chem.*, 269:10991-10995; Stuber et al., 1990, *Immunol. Methods*, 4:121-152, Waeber et al., 1993, *FEBS Letters*, 324:109-112), which can be used to identify and/or purify the fusion protein through binding to a metal ion (Ni) column (QIAEXPRESS Ni-NTA protein Purification System, Quiagen, Inc.). Other affinity tags known in the art may be used and encoded by gene 2.

Insertion of a gene into a plasmid requires that the plasmid be cut at the precise location that the gene is to be inserted. Thus, there must be a restriction endonuclease site at this location (preferably a unique site such that the plasmid will only be cut at a single location during restriction endonuclease digestion). The plasmid is digested, phosphatased, and purified as described above. The gene is then inserted into this linearized plasmid by ligating the two DNAs together. Ligation can be accomplished if the ends of the plasmid are compatible with the ends of the gene to be inserted. If the restriction enzymes are used to cut the plasmid and isolate the gene to be inserted create blunt ends or compatible sticky ends, the DNAs can be ligated together

directly using a ligase such as bacteriophage T4 DNA ligase and incubating the mixture at 16°C for 1-4 hours in the presence of ATP and ligase buffer as described in section 1.68 of Sambrook *et al.*, above. If the ends are not compatible, they must first be made blunt by using the Klenow fragment of DNA polymerase I or bacteriophage T4 DNA polymerase, both of which require the four deoxyribonucleotide triphosphates to fill-in overhanging single-stranded ends of the digested DNA. Alternatively, the ends may be blunted using a nuclease such as nuclease S1 or mung-bean nuclease, both of which function by cutting back the overhanging single strands of DNA. The DNA is then religated using a ligase as described above. In some cases, it may not be possible to blunt the ends of the gene to be inserted, as the reading frame of the coding region will be altered. To overcome this problem, oligonucleotide linkers may be used. The linkers serve as a bridge to connect the plasmid to the gene to be inserted. These linkers can be made synthetically as double stranded or single stranded DNA using standard methods. The linkers have one end that is compatible with the ends of the gene to be inserted; the linkers are first ligated to this gene using ligation methods described above. The other end of the linkers is designed to be compatible with the plasmid for ligation. In designing the linkers, care must be taken to not destroy the reading frame of the gene to be inserted or the reading frame of the gene contained on the plasmid. In some cases, it may be necessary to design the linkers such that they code for part of an amino acid, or such that they code for one or more amino acids.

Between gene 1 and gene 2, DNA encoding a termination codon may be inserted, such termination codons are UAG(amber), UAA (ocher) and UGA (opel). (*Microbiology*, Davis et al. Harper & Row, New York, 1980, pages 237, 245-47 and 274). The termination codon expressed in a wild type host cell results in the synthesis of the gene 1 protein product without the gene 2 protein attached. However, growth in a suppressor host cell results in the synthesis of detectable quantities of fused protein. Such suppressor host cells contain a tRNA modified to insert an amino acid in the termination codon position of the mRNA thereby resulting in production of detectable amounts of the fusion protein. Such suppressor host cells are well known and described, such as *E. coli* suppressor strain (Bullock et al., *BioTechniques* 5:376-379 [1987]). Any acceptable method may be used to place such a termination codon into the mRNA encoding the fusion polypeptide.

The suppressible codon may be inserted between the first gene encoding a polypeptide, and a second gene encoding at least a portion of a phage coat protein. Alternatively, the suppressible termination codon may be inserted adjacent to the fusion site by replacing the last amino acid triplet in the polypeptide or the first amino acid in the phage coat protein. When the plasmid containing the suppressible codon is grown in a suppressor host cell, it results in the detectable production of a fusion polypeptide containing the polypeptide and the coat protein. When the plasmid is grown in a non-suppressor host cell, the polypeptide is synthesized substantially without fusion to the phage coat protein due to termination at the inserted suppressible triplet encoding UAG, UAA, or UGA. In the non-suppressor cell the polypeptide is

synthesized and secreted from the host cell due to the absence of the fused phage coat protein which otherwise anchored it to the host cell.

Gene 1 may encode a mammalian protein, and preferably the protein will be selected from human growth hormone(hGH), N-methionyl human growth hormone, bovine growth hormone, parathyroid hormone, thyroxine, insulin A-chain, insulin B-chain, proinsulin, relaxin A-chain, relaxin B-chain, prorelaxin, glycoprotein hormones such as follicle stimulating hormone(FSH), thyroid stimulating hormone(TSH), leutinizing hormone(LH), glycoprotein hormone receptors, calcitonin, glucagon, factor VIII, an antibody, lung surfactant, urokinase, streptokinase, human tissue-type plasminogen activator (t-PA), bombesin, coagulation cascade factors including factor VII, factor IX, and factor X, thrombin, hemopoietic growth factor, tumor necrosis factor-alpha and -beta, enkephalinase, human serum albumin, mullerian-inhibiting substance, mouse gonadotropin-associated peptide, a microbial protein, such as betalactamase, tissue factor protein, inhibin, activin, vascular endothelial growth factor (VEGF), receptors for hormones or growth factors; integrin, thrombopoietin (TPO), protein A or D, rheumatoid factors, nerve growth factors such as NGF- alpha, platelet-growth factor, transforming growth factors (TGF) such as TGF-alpha and TGF-beta, insulin-like growth factor-I and -II, insulin-like growth factor binding proteins, CD-4, DNase, latency associated peptide, erythropoietin (EPO), osteoinductive factors, interferons such as interferon-alpha, -beta, and -gamma, colony stimulating factors (CSFs) such as M-CSF, GM-CSF, and G-CSF, interleukins (ILs) such as IL-1, IL-2, IL-3, IL-4, IL-6, IL-8, IL-10, IL-12, superoxide dismutase; decay accelerating factor, viral antigen, HIV envelope proteins such as GP120, GP140, atrial natriuretic peptides A, B, or C, immunoglobulins, as well as variants and fragments of any of the above-listed proteins.

The first gene may encode a peptide containing as few as 4-10 amino acid residues and up to about 50 –80 residues. These smaller peptides are useful in determining the antigenic properties of the peptides, in mapping the antigenic sites of proteins, etc. The first gene may also encode a polypeptide of one or more subunits containing more than about 100 amino acid residues which may be folded to form a plurality of rigid secondary structures displaying a plurality of amino acids capable of interacting with the target. Preferably the first gene will be mutated at codons corresponding to only the amino acids capable of interacting with the target so that the integrity of the rigid secondary structures will be preserved.

Phage display of proteins, peptides and mutated variants thereof, including constructing a family of variant replicable vectors containing a transcription regulatory element operably linked to a gene fusion encoding a fusion polypeptide, transforming suitable host cells, culturing the transformed cells to form phage particles which display the fusion polypeptide on the surface of the phage particle, contacting the recombinant phage particles with a target molecule so that at least a portion of the particle bind to the target, separating the particles which bind from those that do not are known and may be used with the transformation method of the invention. See U.S. 5,750,373; WO 97/09446; U.S. 5,514,548; U.S. 5,498,538; U.S. 5,516,637; U.S. 5,432,018; WO 96/22393;

U.S. 5,658,727; U.S. 5,627,024; WO 97/29185; O'Boyle et al, 1997, *Virology*, 236:338-347; Soumillion et al, 1994, *Appl. Biochem. Biotech.*, 47:175-190; O'Neil and Hoess, 1995, *Curr. Opin. Struct. Biol.*, 5:443-449; Makowski, 1993, *Gene*, 128:5-11; Dunn, 1996, *Curr. Opin. Struct. Biol.*, 7:547-553; Choo and Klug, 1995, *Curr. Opin. Struct. Biol.*, 6:431-436; Bradbury and Cattaneo, 1995, *TINS*, 18:242-249; Cortese et al., 1995, *Curr. Opin. Struct. Biol.*, 6:73-80; Allen et al., 1995, *TIBS*, 20:509-516; Lindquist and Naderi, 1995, *FEMS Micro. Rev.*, 17:33-39; Clarkson and Wells, 1994, *Tibtech*, 12:173-184; Barbas, 1993, *Curr. Opin. Biol.*, 4:526-530; McGregor, 1996, *Mol. Biotech.*, 6:155-162; Cortese et al., 1996, *Curr. Opin. Biol.*, 7:616-621; McLafferty et al., 1993, *Gene*, 128:29-36.

In a particularly preferred embodiment, gene 1 encodes the light chain or the heavy chain of an antibody or fragments thereof, such Fab, $F(ab')_2$, Fv, diabodies, linear antibodies, etc. Gene 1 may also encode a single chain antibody (scFv). The preparation of libraries of antibodies or fragments thereof is well known in the art and any of the known methods may be used to construct a family of transformation vectors which may be transformed into host cells using the method of the invention. Libraries of antibody light and heavy chains in phage (Huse et al, 1989, *Science*, 246:1275) and as fusion proteins in phage or phagemid are well known and can be prepared according to known procedures. See Vaughan et al., Barbas et al., Marks et al., Hoogenboom et al., Griffiths et al., de Kruif et al., noted above, and WO 98/05344; WO 98/15833; WO 97/47314; WO 97/44491; WO 97/35196; WO 95/34648; U.S. 5,712.089; U.S. 5,702,892; U.S 5,427,908; U.S. 5,403,484; U.S. 5,432,018; U.S. 5,270,170; WO 92/06176; U.S. 5,702,892. Reviews have also published. Hoogenboom, 1997, *Tibtech*, 15:62-70 ; Neri et al., 1995, *Cell Biophysics*, 27:47; Winter et al., 1994, *Annu. Rev. Immunol.*, 12:433-455; Soderlind et al., 1992, *Immunol. Rev.*, 130:109-124; Jefferies, 1998, *Parasitology*, 14:202-206.

Specific antibodies contemplated as being encoded by gene 1 include antibodies which bind to human leukocyte surface markers, cytokines and cytokine receptors, enzymes, etc. Specific leukocyte surface markers include CD1a-c, CD2, CD2R, CD3-CD10, CD11a-c, CDw12, CD13, CD14, CD15, CD15s, CD16, CD16b, CDw17, CD18-C41, CD42a-d, CD43, CD44, CD44R, CD45, CD45A, CD45B, CD45O, CD46-CD48, CD49a-f, CD50-CD51, CD52, CD53-CD59, CDw60, CD61, CD62E, CD62L, CD62P, CD63, CD64, CDw65, CD66a-e, CD68-CD74, CDw75, CDw76, CD77, CDw78, CD79a-b, CD80-CD83, CDw84, CD85-CD89, CDw90, CD91, CDw92, CD93-CD98, CD99, CD99R, CD100, CDw101, CD102-CD106, CD107a-b, CDw108, CDw109, CD115, CDw116, CD117, CD119, CD120a-b, CD121a-b, CD122, CDw124, CD126-CD129, and CD130. Other antibody binding targets include cytokines and cytokine superfamily receptors, hematopoietic growth factor superfamily receptors and preferably the extracellular domains thereof, which are a group of closely related glycoprotein cell surface receptors that share considerable homology including frequently a WSXWS domain and are generally classified as members of the cytokine receptor superfamily (see e.g. Nicola *et al.*, *Cell*, 67:1-4 (1991) and Skoda, R.C. *et al. EMBO J.* 12:2645-2653 (1993)). Generally, these targets are receptors for

interleukins (IL) or colony-stimulating factors (CSF). Members of the superfamily include, but are not limited to, receptors for: IL-2 (b and g chains) (Hatakeyama *et al.*, *Science*, 244:551-556 (1989); Takeshita *et al.*, *Science*, 257:379-382 (1991)), IL-3 (Itoh *et al.*, *Science*, 247:324-328 (1990); Gorman *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:5459-5463 (1990); Kitamura *et al.*, *Cell*, 66:1165-1174 (1991a); Kitamura *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:5082-5086 (1991b)), IL-4 (Mosley *et al.*, *Cell*, 59:335-348 (1989), IL-5 (Takaki *et al.*, *EMBO J.*, 9:4367-4374 (1990); Tavernier *et al.*, *Cell*, 66:1175-1184 (1991)), IL-6 (Yamasaki *et al.*, *Science*, 241:825-828 (1988); Hibi *et al.*, *Cell*, 63:1149-1157 (1990)), IL-7 (Goodwin *et al.*, *Cell*, 60:941-951 (1990)), IL-9 (Renault *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5690-5694 (1992)), granulocyte-macrophage colony-stimulating factor (GM-CSF) (Gearing *et al.*, *EMBO J.*, 8:3667-3676 (1991); Hayashida *et al.*, *Proc. Natl. Acad. Sci. USA*, 244:9655-9659 (1990)), granulocyte colony-stimulating factor (G-CSF) (Fukunaga *et al.*, *Cell*, 61:341-350 (1990a); Fukunaga *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:8702-8706 (1990b); Larsen *et al.*, *J. Exp. Med.*, 172:1559-1570 (1990)), EPO (D'Andrea *et al.*, *Cell*, 57:277-285 (1989); Jones *et al.*, *Blood*, 76:31-35 (1990)), Leukemia inhibitory factor (LIF) (Gearing *et al.*, *EMBO J.*, 10:2839-2848 (1991)), oncostatin M (OSM) (Rose *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:8641-8645 (1991)) and also receptors for prolactin (Boutin *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:7744-7748 (1988); Edery *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:2112-2116 (1989)), growth hormone (GH) (Leung *et al.*, *Nature*, 330:537-543 (1987)), ciliary neurotrophic factor (CNTF) (Davis *et al.*, *Science*, 253:59-63 (1991) and c-Mpl (M. Souyri *et al.*, *Cell* 63:1137 (1990); I. Vigon *et al.*, *Proc. Natl. Acad. Sci.* 89:5640 (1992)). Still other targets for antibodies made by the invention are erb2, erb3, erb4, IL-10, IL-12, IL-13, IL-15, etc.

Gene 1, encoding the desired polypeptide, may be altered at one or more selected codons. An alteration is defined as a substitution, deletion, or insertion of one or more codons in the gene encoding the polypeptide that results in a change in the amino acid sequence of the polypeptide as compared with the unaltered or native sequence of the same polypeptide. Preferably, the alterations will be by substitution of at least one amino acid with any other amino acid in one or more regions of the molecule. The alterations may be produced by a variety of methods known in the art. These methods include but are not limited to oligonucleotide-mediated mutagenesis and cassette mutagenesis.

Oligonucleotide -mediated mutagenesis is preferred method for preparing substitution, deletion, and insertion variants of gene 1. This technique is well known in the art as described by Zoller et al., *Nucleic Acids Res.*, 10: 6487-6504 (1987). Briefly, gene 1 is altered by hybridizing an oligonucleotide encoding the desired mutation to a DNA template, where the template is the single-stranded form of the plasmid containing the unaltered or native DNA sequence of gene 1. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template will thus incorporate the oligonucleotide primer, and will code for the selected alteration in gene 1.

Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al., *Proc. Nat'l. Acad. Sci. USA*, 75: 5765 (1978).

The DNA template is generated by those vectors that are either derived from bacteriophage M13 vectors (the commercially available M13mp18 and M13mp19 vectors are suitable), or those vectors that contain a single-stranded phage origin of replication as described by Viera et al., *Meth. Enzymol.*, 153: 3 (1987). Thus, the DNA that is to be mutated can be inserted into one of these vectors in order to generate single-stranded template. Production of the single-stranded template is described in sections 4.21-4.41 of Sambrook et al., above.

To alter the native DNA sequence, the oligonucleotide is hybridized to the single stranded template under suitable hybridization conditions. A DNA polymerizing enzyme, usually T7 DNA polymerase or the Klenow fragment of DNA polymerase I, is then added to synthesize the complementary strand of the template using the oligonucleotide as a primer for synthesis. A heteroduplex molecule is thus formed such that one strand of DNA encodes the mutated form of gene 1, and the other strand (the original template) encodes the native, unaltered sequence of gene 1. This heteroduplex molecule is then transformed into a suitable host cell, usually a prokaryote such as *E. coli* JM101. After growing the cells, they are plated onto agarose plates and screened using the oligonucleotide primer radiolabelled with 32-Phosphate to identify the bacterial colonies that contain the mutated DNA.

The method described immediately above may be modified such that a homoduplex molecule is created wherein both strands of the plasmid contain the mutation(s). The modifications are as follows: The single-stranded oligonucleotide is annealed to the single-stranded template as described above. A mixture of three deoxyribonucleotides, deoxyriboadenosine (dATP), deoxyriboguanosine (dGTP), and deoxyribothymidine (dTTP), is combined with a modified thio-deoxyribocytosine called dCTP-(aS) (which can be obtained from Amersham). This mixture is added to the template-oligonucleotide complex. Upon addition of DNA polymerase to this mixture, a strand of DNA identical to the template except for the mutated bases is generated. In addition, this new strand of DNA will contain dCTP-(aS) instead of dCTP, which serves to protect it from restriction endonuclease digestion. After the template strand of the double-stranded heteroduplex is nicked with an appropriate restriction enzyme, the template strand can be digested with ExoIII nuclease or another appropriate nuclease past the region that contains the site(s) to be mutagenized. The reaction is then stopped to leave a molecule that is only partially single-stranded. A complete double-stranded DNA homoduplex is then formed using DNA polymerase in the presence of all four deoxyribonucleotide triphosphates, ATP, and DNA ligase. This

homoduplex molecule can then be transformed into a suitable host cell such as *E. coli* JM101, as described above.

Mutants with more than one amino acid to be substituted may be generated in one of several ways. If the amino acids are located close together in the polypeptide chain, they may be mutated simultaneously using one oligonucleotide that codes for all of the desired amino acid substitutions. If, however, the amino acids are located some distance from each other (separated by more than about ten amino acids), it is more difficult to generate a single oligonucleotide that encodes all of the desired changes. Instead, one of two alternative methods may be employed.

In the first method, a separate oligonucleotide is generated for each amino acid to be substituted. The oligonucleotides are then annealed to the single-stranded template DNA simultaneously, and the second strand of DNA that is synthesized from the template will encode all of the desired amino acid substitutions. The alternative method involves two or more rounds of mutagenesis to produce the desired mutant. The first round is as described for the single mutants: wild-type DNA is used for the template, an oligonucleotide encoding the first desired amino acid substitution(s) is annealed to this template, and the heteroduplex DNA molecule is then generated. The second round of mutagenesis utilizes the mutated DNA produced in the first round of mutagenesis as the template. Thus, this template already contains one or more mutations. The oligonucleotide encoding the additional desired amino acid substitution(s) is then annealed to this template, and the resulting strand of DNA now encodes mutations from both the first and second rounds of mutagenesis. This resultant DNA can be used as a template in a third round of mutagenesis, and so on.

Cassette mutagenesis is also a preferred method for preparing substitution, deletion, and insertion variants of gene 1. The method is based on that described by Wells et al., *Gene*, 34:315 (1985). The starting material is the plasmid (or other vector) comprising gene 1, the gene to be mutated. The codon(s) in gene 1 to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in gene 1. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are compatible with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated DNA sequence of gene 1.

In a preferred embodiment, gene 1 is linked to gene 2 encoding at least a portion of a phage coat protein. Preferred coat protein genes are the genes encoding coat protein 3 and coat protein 8 of filamentous phage specific for *E. coli*, such as M13, f1 and fd phage. Transfection of

host cells containing a replicable expression vector which encodes the gene fusion of gene 1 and gene 2 and production of phage particles according to standard procedures provides phage particles in which the polypeptide encoded by gene 1 is displayed on the surface of the phage particle.

Although published protocols suggest using a final cell concentration of about 10^{10} colony forming units (cfu)/mL, the present invention allows one to obtain cell concentrations of 5×10^{10} cfu/mL of viable living cells and greater for use in electroporation. Preferably, the viable cells are concentrated to about 1×10^{11} to about 4×10^{11} cfu/mL in the method of the invention. Preferred cells which may be concentrated to this range are the SS320 cells described below. Although Dower et al. indicate that the yield of transformants should increase with the number of cells present during electroporation, it is believed that cell concentrations above about 5×10^{10} cells/mL have not been used in practice. It has now been discovered that some cells, in particular, *E. coli* strains, can be concentrated to concentrations far greater than has been previously suggested. A crucial factor in determining the maximum final concentration of a given strain is the resistance of the strain to the standard washing steps used in the preparation of electrocompetent cells. It has been discovered that the proportion of cells surviving the washing procedure varies. As a part of this invention, it has been discovered that prior methods of preparing cells for electroporation often result in higher numbers of non-viable cells and lower transformation yields. In this embodiment, cells are grown in culture in standard culture broth, optionally for about 6-48 hrs (or to $OD_{600} = 0.6 - 0.8$) at about 37°C, and then the broth is centrifuged and the supernatant removed (e.g. decanted). Initial purification is preferably by resuspending the cell pellet in a buffer solution (e.g. HEPES pH 7.4) followed by recentrifugation and removal of supernatant. The resulting cell pellet is resuspended in dilute glycerol (e.g. 5 – 20% v/v) and again recentrifuged to form a cell pellet and the supernatant removed. The final cell concentration is obtained by resuspending the cell pellet in water or dilute glycerol to the desired concentration. As noted above, it has been discovered that these washing steps have an effect on cell survival, that is on the number of viable cells in the concentrated cell solution used for electroporation. It is preferred to use cells which survive the washing and centrifugation steps in a high survival ratio relative to the number of starting cells prior to washing. Most preferably, the ratio of the number of viable cells after washing to the number of viable cells prior to washing is 1.0, i.e., there is no cell death. However, the survival ratio may be about 0.8 or greater, preferably about 0.9 – 1.0.

A particularly preferred recipient cell is the electroporation competent *E. coli* strain of the present invention, which is *E. coli* strain MC1061 containing a phage F′ episome. Any F′ episome which enables phage replication in the strain may be used in the invention. Suitable episomes are available from strains deposited with ATCC or are commercially available (CJ236, CSH18, DH5alphaF', JM101, JM103, JM105, JM107, JM109, JM110), KS1000, XL1-BLUE, 71-18 and others). Strain SS320 was prepared by mating MC1061 cells with XL1-BLUE cells under conditions sufficient to transfer the fertility episome (F′ plasmid) of XL1-BLUE into the MC1061

cells. In general, mixing cultures of the two cell types and growing the mixture in culture medium for about one hour at 37°C is sufficient to allow mating and episome transfer to occur. The new resulting *E. coli* strain has the genotype of MC1061 which carries a streptomycin resistance chromosomal marker and the genotype of the F' plasmid which confers tetracycline resistance. The progeny of this mating is resistant to both antibiotics and can be selectively grown in the presence of streptomycin and tetracycline. Strain SS320 has been deposited with the American Type Culture Collection (ATCC), 10801 University Boulevard, Manassas, Virginia, USA on June 18, 1998 and assigned Deposit Accession No. 98795.

This deposit of strain SS320 was made under the provisions of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purpose of Patent Procedure and the Regulations thereunder (Budapest Treaty). This assures maintenance of a viable culture for 30 years from the date of deposit. The organisms will be made available by ATCC under the terms of the Budapest Treaty, and subject to an agreement between Genentech, Inc. and ATCC, which assures permanent and unrestricted availability of the progeny of the cultures to the public upon issuance of the pertinent U.S. patent or upon laying open to the public of any U.S. or foreign patent application, whichever comes first, and assures availability of the progeny to one determined by the U.S. Commissioner of Patents and Trademarks to be entitled thereto according to 35 USC §122 and the Commissioner's rules pursuant thereto (including 37 CFR §1.14 with particular reference to 886 OG 638).

The assignee of the present application has agreed that if the cultures on deposit should die or be lost or destroyed when cultivated under suitable conditions, they will be promptly replaced on notification with a viable specimen of the same culture. Availability of the deposited cultures is not to be construed as a license to practice the invention in contravention of the rights granted under the authority of any government in accordance with its patent laws.

SS320 cells have properties which are particularly favorable for electroporation. It has been discovered that SS320 cells are particularly robust and are able to survive multiple washing steps with higher cell viability than most other electroporation competent cells. The ability of SS320 cells to survive washing steps allows one to prepare cell concentrations which are greater than the cell concentrations suggested by the Dower et al. protocol. Other strains suitable for use with the higher cell concentrations include TB1, MC1061, etc. These higher cell concentrations provide greater transformation efficiency for the process of the invention.

The use of higher DNA concentrations during electroporation (about 10X) increases the transformation efficiency and increases the amount of DNA transformed into the host cells. The use of higher cell concentrations also increases the efficiency (about 10X). The larger amount of transferred DNA produces larger libraries having greater diversity and representing a greater number of unique members of a combinatorial library. The method of the invention is useful to increase the size of an expressible combinatorial library by a factor of about 100X with a single

electroporation event which allows one to select, amplify and identify rare library members present in amounts 100X lower than with comparable conventional methods.

Dower et al. have demonstrated that saturation (transformation of most survivors of an electroporation) can be achieved with closed circular DNA concentrations of about 10 microgram/mL. However, the construction of libraries, for example a library of fusion genes encoding fusion polypeptides, necessarily involves the introduction of DNA fragments representing the library into a suitable vector to provide a family or library of vectors. In the case of cassette mutagenesis, the synthetic DNA is a double stranded cassette while in fill-in mutagenesis the synthetic DNA is single stranded DNA. In either case, the synthetic DNA is incorporated into a vector to yield a reaction product containing closed circular double stranded DNA which can be transformed into a cell to produce the library. However, processes used to incorporate synthetic DNA into closed circular DNA are generally less than 100% efficient, and often the desired closed circular product represents only a small fraction of the total DNA. To achieve saturation with a ligation or fill-in reaction product may require significantly higher DNA concentrations than those necessary when using pure closed circular DNA. Prior methods do not allow or suggest using DNA concentrations which are sufficient to achieve saturation in the electroporation reactions using DNA incorporating synthetic DNA fragments into transformation vectors, e.g. plasmids, phage vectors, phagemid vectors, etc.

For example, the reaction of Example 6 below, demonstrates that using the method of the invention, a DNA concentration of 19 microgram/mL results in transformation of 53% of surviving cells. In this example, the fill-in reaction ws very efficient yielding about 95% of the desired closed circular DNA product as evidenced by agarose gel electrophoresis. The DNA concentration can be increased to hundreds (e.g., 300 – 500 microgram/mL) of micrograms/mL without adversely affecting either the cell survival or the transformation. The maximum number of transformants is obtained at a DNA concentration 20-fold lower than the maximum DNA concentration tested and even the largest DNA concentration tested had no detrimental effect on the electroporation reaction. The invention provides a useable dynamic range of DNA concentration in electroporation which far exceeds that possible with prior methods. Using the method of the invention, inefficient transformation vector formation reactions which yield only small amounts of the desired clonable DNA, for example only 10-50%, or even 1-10%, can be made to saturate the electroporation survivors by using DNA concentrations up to about 400 -500 micrograms/mL.

The method of the invention also allows the facile introduction of two or more vectors into a single cell, even with a DNA reaction preparation in which the desired vector (e.g. closed circular DNA) represents only a fraction of the total DNA in the preparation. This makes possible the simultaneous introduction of multiple foreign genes in separate vectors, for example, two or more libraries into a single transformant. The introduction multiple members of a library into a single cell expands the library diversity beyond the number of transformants. For example, if on average

a transformant in an electroporation reaction maintains two plasmids, the library diversity will be twice the number of transformants.

The saturation concentration for a given DNA preparation can be defined as the concentration beyond which further increases in DNA concentration do not result in increased transformation yields.

At the saturation concentration, all cells capable of taking up and maintaining DNA have been transformed. Dower et al., above, have shown that transformation efficiency is directly proportional to DNA concentration. An increase in DNA concentration should result in corresponding increases in the average number of unique plasmids per survivor. For example, if at a given DNA concentration, electroporation results in survivors carrying one unique plasmid each on average, then doubling the DNA concentration will result in the survivors carrying two unique plasmids each on average.

The current invention allows for the use of DNA concentrations in electroporation at least an order of magnitude above that possible with prior art methods. This in turn allows for the simultaneous introduction of two or more unique plasmids (containing different library members) into a single cell during a single electroporation reaction, even with DNA preparations which have been enzymatically manipulated.

Several members of a single library can be introduced into a single cell and thus the library diversity can be expanded beyond the number of transformants. For example, if on average a transformant produced in a given electroporation reaction maintains two plasmids, the library diversity will be twice the number of transformants. In the case of phage display libraries, packaging of phage DNA or phagemids maintained within the same transformant will result in a random display of the different fusion proteins produced from the different phage DNA or phagemids. Thus, some of the fusion proteins will be displayed in association with their cognate DNA sequences while others will be associated with completely unrelated sequences which happened to co-transform by random chance.

In the case of highly polyvalent display (e.g., peptide display on protein-8) where the number of incorporated fusion proteins per phage particle greatly exceeds the number of co-transformed phage DNA or phagemids, each phage particle will display its cognate fusion protein along with other unrelated fusion proteins. This will result in an increase in background during the first round of sorting due to the capture of fusion proteins associated with phage containing unrelated DNA. If the phage captured in the first round are amplified by infection into *E. coli* at a low multiplicity of infection so that each *E. coli* cell is infected by only one phage, the correlation between displayed fusion protein and its DNA sequence will be restored. The second round of sorting will then eliminate DNA sequences which do not encode proteins with affinity for the target.

In the case of monovalent display (e.g., protein display on protein-3) where the ratio of fusion protein to phage is less than one, each displayed fusion protein may associate either with its

cognate DNA sequence or with an unrelated co-transformed DNA sequence. Provided that the number of phage used in the first sort is large enough to ensure that at least some DNA sequences are linked with the fusion proteins they encode, phage displaying fusion proteins with affinity for a given target (and containing the cognate DNA sequence) will be selectable. As in the case of polyvalent display, incorrect linkage between fusion protein and phage DNA will result in the first round capture of incorrect DNA sequences, but these sequences can be eliminated in the second round as described above.

The use of DNA concentrations far above the saturation concentration also allows for the co-transformation of a cell with plasmids from different libraries in a single electroporation. Thus the current invention can be used to facilitate and simplify any methodology which requires that a single cell maintain two or more plasmids from distinct libraries, since two or more libraries can be introduced simultaneously rather than serially.

For example, Griffiths et al. (*EMBO Journal* 13(14):3245-3260, 1994) used the process of combinatorial infection and *in vivo* recombination to increase the size of a phage antibody repertoire. The process involved the separate electroporation of light chain and heavy chain repertoires to produce two distinct libraries; a third step combined the two libraries to yield the desired antibody library. The DNA concentrations achievable with the current invention enables the co-transformation of the two libraries in a single electroporation.

The transformed cells are generally selected by growth on an antibiotic, commonly tetracycline (tet) or ampicillin (amp), to which they are rendered resistant due to the presence of tet and/or amp resistance genes in the vector.

Suitable phage and phagemid vectors for use in this invention include all known vectors for phage display. Additional examples include pComb8 (Gram, H., Marconi, L. A., Barbas, C. F., Collet, T. A., Lerner, R. A., and Kang, A.S. (1992) *Proc. Natl. Acad. Sci. USA* 89:3576-3580); pC89 (Felici, F., Catagnoli, L., Musacchio, A., Jappelli, R., and Cesareni, G. (1991) *J. Mol. Biol.* 222:310-310); pIF4 (Bianchi, E., Folgori, A., Wallace, A., Nicotra, M., Acali, S., Phalipon, A., Barbato, G., Bazzo, R., Cortese, R., Felici, F., and Pessi, A. (1995) *J. Mol. Biol.* 247:154-160); PM48, PM52, and PM54 (Iannolo, G., Minenkova, O., Petruzzelli, R., and Cesareni, G. (1995) *J. Mol. Biol* ,248:835-844); fdH (Greenwood, J., Willis, A. E., and Perham, R. N. (1991) *J. Mol. Biol,*. 220:821-827); pfd8SHU, pfd8SU, pfd8SY, and fdISPLAY8 (Malik, P. and Perham, R. N. (1996) *Gene,* 171:49-51); "88" (Smith, G. P. (1993) *Gene,* 128:1-2); f88.4 (Zhong, G., Smith, G. P., Berry, J. and Brunham, R. C. (1994) *J. Biol. Chem,* 269:24183-24188); p8V5 (Affymax); MB1, MB20, MB26, MB27, MB28, MB42, MB48, MB49, MB56: Markland, W., Roberts, B. L., Saxena, M. J., Guterman, S. K., and Ladner, R. C. (1991) *Gene,* 109:13-19). Similarly, any known helper phage may be used when a phagemid vector is employed in the phage display system. Examples of suitable helper phage include M13-KO7 (Pharmacia), M13-VCS (Stratagene), and R408 (Stratagene).

After selection of the transformed cells, these cells are grown in culture and the vector DNA may then be isolated. Phage or phagemid vector DNA can be isolated using methods known in the art, for example, as described in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd edition, (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

The isolated DNA can be purified by methods known in the art such as that described in section 1.40 of Sambrook et al., above and as described above. This purified DNA can then be analyzed by DNA sequencing. DNA sequencing may be performed by the method of Messing et al., *Nucleic Acids Res.*, 9:309 (1981), the method of Maxam et al., *Meth. Enzymol.*, 65:499 (1980), or by any other known method.

The invention also contemplates producing product polypeptides which have been obtained by culturing a host cell transformed with a replicable expression vector, where the replicable expression vector contains DNA encoding a product polypeptide operably linked to a control sequence capable of effecting expression of the product polypeptide in the host cell; where the DNA encoding the product polypeptide has been obtained by:

(a) constructing a family of variant replicable plasmids containing a transcription regulatory element operably linked to a gene fusion encoding a fusion protein, wherein the gene fusion contains a first gene encoding a polypeptide and a second gene encoding at least a portion of a phage coat protein, where the variant replicable plasmids contain variant first genes encoding variant polypeptides;

(b) transforming suitable host cells with the plasmids using the method of the invention;

(c) optionally, when the plasmid is a phagemid which requires a helper phage to produce phage particles, infecting the transformed host cells with an amount of helper phage encoding the phage coat protein sufficient to produce recombinant phagemid particles which display the fusion protein on the surface of the particles, preferably where no more than a minor amount of the phagemid particles display one or more copies of the fusion protein on the surface of the phagemid particles;

(d) culturing the transformed infected host cells under conditions suitable for forming recombinant phage particles containing at least a portion of the plasmid and capable of transforming the host cells;

(e) contacting the recombinant phage particles with a target molecule so that at least a portion of the phage particles bind to the target molecule;

(f) separating phage particles that bind to the target molecule from those that do not bind;

(g) selecting one of the variant polypeptides encoded in a phage particle which binds or does not bind to the target molecule as the product polypeptide and cloning DNA encoding the product polypeptide into the replicable expression vector; and recovering the expressed product polypeptide; and product polypeptides produced by the process.

U.S. 5,750,373 describes generally how to produce and recover a product polypeptide by culturing a host cell transformed with a replicable expression vector (e.g., a phagemid) where the

DNA encoding the polypeptide has been obtained by steps (a)-(f) above using conventional helper phage where a minor amount (<20%, preferably <10%, more preferably < 1%) of the phage particles display the fusion protein on the surface of the particle. Any suitable helper phage may be used to produce recombinant phagemid particles, e.g., VCS, etc. The present invention provides an improved method by transforming the host cells by electroporation using the high DNA concentrations and other embodiments of the invention. One of the variant polypeptides obtained by the phage display process may be selected for larger scale production by recombinant expression in a host cell. Culturing of a host cell transformed with a replicable expression vector which contains DNA encoding a product polypeptide which is the selected variant operably linked to a control sequence capable of effecting expression of the product polypeptide in the host cell and then recovering the product polypeptide using known methods is part of this invention.

B. Novel coat proteins, coat fusion proteins, vectors, cells and methods

The expression of polypeptides on the surface of bacteriophage has been developed and refined over several years. In particular, systems have been developed for displaying recombinant peptides, proteins, antigens and antibodies on the surface of filamentous bacteriophage. A number of filamentous phage have been identified which are able to infect gram negative bacteria, such as *E. coli*. These phage have a single stranded covalently closed DNA genome containing only about 10 genes encased in a cylinder of coat proteins. Due to the relative simplicity of these viruses and the ease with which they can be genetically manipulated, filamentous phage have been well studied. All strains of filamentous phage have a similar virion structure and life cycle. Upon infection, viral DNA enters the cell and is converted to a double stranded replicative form by host enzymes. Progeny DNA is replicated by a rolling circle mechanism and is assembled with a viral replication assembly protein into an elongated DNA/protein complex. The virion is extruded through the membrane of the host cell where the replication assembly protein is replaced by coat proteins. The virion sheath contains several thousand identical α-helical proteins as the major coat protein.

Foreign DNA can be inserted as a separate gene in a viral intergenic region. When the heterologous DNA is inserted as a separate gene, the virus or a virus-derived plasmid (phagemid) becomes a cloning vector. When the heterologous DNA is inserted as a gene fusion with a coat protein of the virus, the virus or phagemid is capable of displaying the polypeptide encoded by the heterologous DNA as a fusion protein on the surface of the virion. Fusion proteins containing variants of the major coat protein of any bacteriophage which is suitable for use in a known phage display system are within the scope of the present invention. Class I and class II filamentous phage are included within the scope of the invention. Class I includes strains Ff, IKe and If1; class II includes strains Pf1, Pf3 and Xf. The Ff phage include the virtually identical strains fd, f1 and M13.

The structure and function of the major coat protein of filamentous bacteriophage have been studied in order to understand the interactions between phage DNA and the coat proteins, as well as to understand the forces which effect packing of the coat proteins into bacteriophage particles. Point mutations in the major coat proteins of filamentous bacteriophage have been prepared to assist in these studies. Hunter, E.J. et al., (1987) *Nature*, 327:252; Greenwood, J. et al., (1991) *J. Mol. Biol.* 217: 223; Deber, C.M. et al., (1993) *Proc. Natl. Acad. Sci. USA*, 90:11648; Symmons, M.S. et al., (1995) *J. Mol. Biol.* 245:86; Williams, K.A. et al., (1995) *J. Mol. Biol.* 252:6; Spruijt, R. B. et al., (1996) *Biochemistry* 35:10383; Marvin, D.A. (1998) *Current Opinion in Structural Biology* 8:150; Haigh, N.G. and Webster, R.E., (1998) *J. Mol. Biol.*, 279:19. These studies suggest that some point mutations are tolerated by phage and result in packaging of phage particles containing the mutant major coat proteins. None of these studies involve fusion proteins of heterologeous polypeptides to variant phage coat proteins, however. Furthermore, it is known that the inclusion of fusion proteins in a phage coat may hinder phage packaging giving rise to poor phage yields and/or may prevent display of the fusion protein on the surface of the phage even when a wild type coat protein sequence is used (Smith, G. P. (1985), *Science*, 228:1315). Whereas small peptides (10 - 15 amino acid residues) can generally be displayed in up to about 800 - 1000 copies per virion, full length proteins are displayed in many fewer numbers (1 - 10 copies per virion). Malik, P. et al. (1996) J. Mol. Biol. 260:9.

The sequences of several known mature major coat proteins of filamentous bacteriophage aligned with the mature M13 coat protein VIII are shown in the Table below. Segments of the coat proteins were aligned with M13 protein VIII so as to provide maximum identity with the M13 protein without the introduction of any deletions or insertions. Numbering above the sequences refers to the residues of mature M13 protein VIII. Protein sequences are taken from the Dayhoff protein database (accession numbers: M13, COAB_BPFD; F1, COAB_BPFD; Fd, COAB_BPFD; Zj-2, COAB_BPZJ2; If-1, COAT_BPIF1; I2-2, COAB_BPI22; Ike, COAB_BPIKE). Homologous residues are indicated with dashes. A sequence having a single deletion is also known (WO 92/18619). It can be seen that there is considerable homology among the sequences of these coat proteins, particularly among the M13, f1, fd and Zj-2 coat proteins and among the If1, I22 and Ike coat proteins.

Table

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
M13	A	E	G	D	D	P	A	K	A	A	F	N	S	L	Q	A	S	A	T	E	Y	I	G	Y	A
F1	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-
Fd	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-
Zj-2	-	-	-	-	-	-	-	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	-	-	-
Ifl	D	D	A	T	S	Q	-	-	-	-	-	D	-	-	T	-	Q	-	-	-	M	S	-	-	-
I2-2	S	T	A	T	S	Y	-	T	E	-	M	-	-	-	K	T	Q	-	-	D	L	-	D	Q	T
Ike	N	A	A	T	N	Y	-	T	E	-	M	D	-	-	K	T	Q	-	I	D	L	-	S	Q	T

	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50
M13	W	A	M	V	V	V	I	V	G	A	T	I	G	I	K	L	F	K	K	F	T	S	K	A	S
F1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fd	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-
Zj-2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	R	-	-
Ifl	-	-	L	-	-	L	V	-	-	-	-	V	-	-	-	-	-	-	-	-	S	-	-	-	V
I2-2	-	P	V	-	T	S	V	A	V	-	G	L	A	-	R	-	-	-	-	-	S	-	-	-	V
Ike	-	P	V	-	T	T	V	-	V	-	G	L	V	-	R	-	-	-	-	-	S	-	-	-	V

(SEQ ID NOS. 2 - 8)

Filamentous phage particle are formed when the phage genes are transcribed, translated and replicated in a host cell. Phage coat proteins are directed to the periplasm and temporarily lodge in the cell membrane with a portion of the coat protein in the periplasm (the periplasmic domain), a portion of the coat protein in the cytoplasm (the cytoplasmic domain), and a portion of the coat protein spanning the cell membrane (the transmembrane domain). Phage particles are formed when the coat proteins assemble around the phage DNA as the phage particle passes through the cell membrane. The M13 major coat protein contains 50 residues which can be divided into three regions: The periplasmic domain contains residues 1 to 20, the transmembrane domain contains residues 21 to 39, and the cytoplasmic domain contains residues 40 to 50 (Marvin, D.A. (1998) *Current Opinion in Structural Biology* 8:150). The other major coat proteins in the table above have a similar domain structure.

Surprisingly, applicants have also discovered that fusion proteins of heterologeous polypeptides to variants of the major coat proteins of bacteriophage are well tolerated in phage display systems. This result was unexpected since previous phage display systems have used the wild type coat protein sequences, generally of M13 or fragments thereof, T4, T7 or lambda phage coat proteins. In one aspect of the present invention, phage display and selection have been used to obtain bacteriophage displaying fusion proteins on the surface thereof where the fusion protein is a heterologeous polypeptide fused to a phage major coat protein variant having one or more amino acid substitutions, deletions or additions. Fusion proteins having a heterologous polypeptide linked to a variant of the coat proteins in the Table above are within the scope of this invention.

Preferred variants of M13, f1 and fd coat protein VIII contain at least one amino acid residue selected from the lists below in the position indicated:

Residue Number	Sample Substitutions	Preferred Amino Acids
1	E, L, V, Q (neg. charged, hydrophobic, polar)	D, I, N
2	R, H, F, W, E (charged, large aromatic)	K, Y, D
3	T, E, L (small polar, neg. charged, hydrophobic)	S, D, I, V, A
4	D, R, H (charged)	E, K
5	R, H, N, D (charged, polar)	K, Q, E
6	Y, W, S, I, L (aromatic, small polar, hydrophobic)	F, T, V
7	T, N (small polar)	S
8	D, H (charged)	R, E, K
9	E, Q, T (neg. charged, polar)	D, N, S
11	W, I, V (aromatic, hydrophobic)	Y, L, F
12	R, H, N (charged, polar)	E, D, K, Q
13	I, L, E, Q (neg. charged, polar)	A, V, D, T, N, S
14	L (hydrophobic)	I, V

15	D, R, N (charged, polar)	E, K, H, Q
16	E, V, L, T (neg. charged, hydrophobic, polar)	D, I, A, S, G
17	E, V, L (neg. charged, hydrophobic, polar)	I, A, T, D
18	L (hydrophobic)	V, I
19	L, T, Q, E (hydrophobic, polar, neg. charged)	I, V, S, A, N, D
20	R, D, H (charged)	N, Q, K, E
21	W, Y, I (aromatic, hydrophobic)	L, F, V
22	W (aromatic)	F, Y
23	W, Y, I, V, H, K (aromatic, hydrophobic, pos. charged)	F, L, R
24	I, Q (hydrophobic, polar)	L, N, V
25	S (hydrophobic, polar)	L, I, T, V
26	A, I, V (hydrophobic)	G, L, M
27	N (small polar)	T, S
28	I, L (hydrophobic)	V
29	K, R, F, W (aromatic, pos. charged)	H, Y
30	I, V (hydrophobic)	L

In these tables, the letter code refers to amino acid residues as follows: A (Ala) alanine; B (Asx) asparagine or aspartic acid; C (Cys) cysteine; D (Asp) aspartic acid; E (glu) glutamic acid; F (Phe) phenylalanine; G (Gly) glycine; H (His) histidine; I (Ile) isoleucine; K (Lys) lysine; L (Leu) leucine; M (Met) methionine; N (Asn) asparagine; O (Xaa) stop codon; P (Pro) proline; Q (Gln) glutamine; R (Arg) arginine; S (Ser) serine; T (Thr) threonine; V (Val) valine; W (Trp) tryptophan; X (Xaa) unknown or non-standard; Y (Tyr) tyrosine; Z (Glx) gluamine or glutamic acid.

As a part of this invention, it has been discovered that the amino acid sequence of phage major coat proteins can be modified to produce variants of the major coat protein which are useful as components of fusion proteins in phage display systems and methods. Fusion proteins containing variants of the major coat protein of a bacteriophage influence the ability of phage to package the fusion proteins into complete virus particles (virions). That is, variants of the major coat proteins can be used to alter the number of fusion proteins incorporated into a virus particle. Hyper-functional variants of the major coat protein can be used to increase the number of fusion proteins incorporated into a virus particle. Conversely, hypo-functional variants can be used to decrease fusion protein incorporation. In this way, the present invention provides a method for tailoring the incorporation of fusion proteins into virus particles to achieve a desired level of valency. This is particularly important for fusion proteins in which the heterologous polypeptide is relatively large, for example, where the heterologous polypeptide contains 50 or more amino acids, preferably 100 or more amino acids, and even more preferably 200 or more amino acid residues and also where the heterologous polypeptide is a protein having secondary and tertiary structure. The method of the invention, therefore, provides a means of overcoming the deficiencies of prior

art phage display methods which utilize the major coat protein of a bacteriophage and which generally obtain only limited incorporation of the fusion protein into the virus coat. The fusion polypeptides of the invention are able to function in known phage display systems by substituting for the conventionally used wild type coat protein fusions with heterologous polypeptides. The fusion polypeptides of the invention will function in a similar manner to conventional fusion proteins in each of the known phage display systems, in which the fusion is with the major coat protein of the virus, further allowing one to select the degree of valency or number of fusion proteins displayed on the surface of the phage with more reliability. For example, the phage and phagemid vectors and the phage display systems described in U.S. 5,223,409; U.S. 5,403,484; U.S. 5,571,689; U.S. 5,750,373, and U.S. 5,780,279 (and others noted above) can be modified to use the fusion proteins of the invention to improve display of peptides, proteins, antibodies and fragments thereof on the surface of phage. The phage is preferably a DNA phage.

In addition to filamentous phage, the invention is suitable for use in phage display systems using lambda phage, Baculovirus, T4 phage and T7 phage. In each of these display systems, the coat protein used to display a heterologous polypeptide is mutated to form variants of the coat protein using the method of the invention and variants having the desired degree of display (hyper-functional or hypo-functional variants) are selected. The selected variant coat protein is then used to form a fusion protein with a heterologous polypeptide which is to be displayed on the surface of the virus particles. The scope of this invention includes the method(s) of the invention using these phage as well as fusion proteins, replicable expression vectors containing a gene encoding the fusion protein, virus particles containing the fusion proteins or vectors, host cells containing the virus particles, fusion proteins or vectors, libraries containing a plurality of different individuals of these fusion proteins, vectors, virions, cells, etc.

Polypeptides may be displayed on lambdoid phage using coat proteins in either the head or the tail portions of the phage particle (U.S. 5,627,024). Suitable head proteins include proteins pE, pD, pB, pW, pFII, pB* (a cleavage product of pB), pXI, and pX.2; suitable tail proteins include pJ, pV, pG, pM, and pT. The structure and location of these coat proteins is well known. See Georgeopoulos, et al. and Katsura in "Lambda II", R. W. Hendrix et al. eds. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983. Preferred lambda proteins for use in the invention are the tail coat proteins, particularly pV. U.S. 5,627,024 describes how to display polypeptides on lambda phage, preferably using pV. The fusion proteins of the invention, therefore, include at least a portion of variants of pE, pD, pB, pW, pFII, pB*, pXI, pX.2, pJ, pV, pG, pM, and pT fused to a heterologous polypeptide.

Polypeptides can also be displayed on T4 phage. The structure of the T4 virion is well studied. See Eiserling in "Bacteriophage T4", C. K. Mathews et al. eds. American Society for Microbiology, Washington, D.C., 1983, pp 11-24. Peptides and full length proteins may be displayed as fusions with the SOC (small outer capsid protein) and the HOC (highly antigenic outer capsid protein) coat proteins of T4 phage. Further, the minor T4 fibrous protein fibritin

encoded by the wax (whisker's antigen control) gene can be lengthened at the C terminus with a heterologous polypeptide to form a fusion protein which is displayed on the T4 whisker protein. See Ren, Z-J. et al. (1998) Gene 215:439; Zhu, Z. (1997) CAN 33:534; Jiang, J et al. (1997) can 128:44380; Ren, Z-J. et al. (1997) CAN 127:215644; Ren, Z-J. (1996) Protein Sci. 5:1833; and Efimov, V. P. et al. (1995) Virus Genes 10:173.

T7 phage may also be used to display polypeptides and proteins. Smith, G. P. and Scott, J.K. (1993) Methods in Enzymology, 217, 228-257; U.S. 5,766,905. Commercial kits (T7Select1-1 and T7Select415-1 from Novagen) are available for display of polypeptides as fusion proteins with the 10B capsid protein (397 amino acids) and with the 10A capsid protein (344 amino acids). These systems are easy to use and have the capacity to display peptides up to about 50 amino acids in size in high copy number (415 per phage), and proteins up to about 1200 amino acids in low copy number (0.1-1 per phage). T7 is a double stranded DNA phage that has been extensively studied (Dunn, J.J. and Studier, F.W. (1983) J. Mol. Biol. 166:477-535; Steven, A.C. and Trus, B.L. (1986) Electron Microscopy of Proteins 5:1-35). Phage assembly takes place inside the host (E. coli) cell and mature phage are released by cell lysis. Fusion proteins of heterologous polypeptides to variants of T7 coat proteins, such as 10B and 10A, vectors containing a gene encoding the fusion protein, etc. are within the of the invention. Preferably, fusion proteins are prepared by altering, preferably by mutating to a non-wild type amino acid, one or more of residues 1-348 of capsid protein 10B.

The invention also includes fusion proteins of heterologous polypeptides with Baculovirus coat protein variants. Baculovirus expression vectors, particularly those based on Autographa californica nuclear polyhedrosis virus, are easily generated and are now widely used for the expression of heterologous polypeptides in cultured insect cells and insect larvae (Weyer, U. and Possee, R. D. (1991) *J. Gen. Virol.* 72:2967). These viruses contain a double stranded, circular genome, where foreign genes can be inserted easily. Tarui, H. et al. (1995) *J. Fac. Agr. Kyushu Univ.*, 40;45. It is possible to display a glycosylated eukaryotic protein on the surface of baculovirus particles, using a fusion with the baculovirus major coat protein gp64 or at least by fusing the heterologous polypeptides to the membrane anchorage domain of gp64 only. The efficiency of various promoters (polyhedrin, basic, gp64-promoter) have been examined, including the "very late" polyhedrin promoter and the "early and late" gp64 promoter. In order to express a foreign gene on the surface of baculoviruses efficiently, it is necessary to choose a regulating promoter, that on one hand will transcribe sufficient amounts of the target protein, and on the other hand start transcription early enough in the viral replication cycle, to guarantee efficient packaging, complete glycosylation and correct folding.

As a further aspect of the invention, it has been discovered that phage display technology can be applied to the major coat protein itself to generate useful major coat protein variants and fusion proteins thereof. In this aspect of the invention, a library of replicable expression vectors is constructed where the expression vector includes a transcription regulatory element operably

linked to a gene fusion encoding a fusion protein where the gene fusion contains a first gene encoding a first polypeptide and a second gene encoding a variant of the major coat protein of the bacteriophage used in the phage display system. That is, a library is constructed in which the second gene encodes for a plurality of variant phage major coat proteins and the phage display system or method is used to select for the variant sequence or sequences which give the desired degree of fusion protein surface display of a polypeptide. The degree of randomization of the major coat protein to produce the variants is optional. That is, libraries may be constructed in which each amino acid residue in the coat protein may be randomized to any amino acid or each residue may be limited to a subset of amino acids to produce a more limited library having predetermined constraints. It is also possible to construct a library in which a subset of residues are allowed to vary within a subset of amino acids, i.e. selected residues are incompletely randomized. For example, it is possible to limit the range of variants for a particular amino acid residue to polar amino acids, hydrophobic amino acids, hydrophilic amino acids, aromatic amino acids, positively or negatively charged amino acids, sterically small or large amino acids, or to a particular desired combination of amino acids to obtain a smaller library having particular constraints. Any combination of amino acids may be used to prepare the desired libraries.

It is also possible to produce libraries in which amino acids residues within desired segments of the major coat protein are varied to obtain a library of major coat protein variants having amino acid additions, substitutions or deletions within defined regions of the coat protein. As an example, the major coat protein may be divided into an arbitrary number of zones, generally 2-10 zones, and a library constructed of variants within one or more of the zones. The mature major coat proteins of M13, f1 and fd phage, for example, contain 50 amino acids and might be divided into 10 zones of 5 amino acid residues each or into zones with unequal numbers of residues in each zone, e.g. zones containing 15, 10, 9, and 8 residues. Zones corresponding to the cytoplasmic, transmembrane and periplasmic regions of the coat protein may be used. A separate library may be constructed for each of the zones in which amino acid alterations are desired. If fusion proteins are desired in which the major coat protein variant has an amino acid alteration in zone 1, for example, a single library may be constructed in which one or more of the amino acid residues within zone 1 is varied. Alternatively, one may wish to produce fusion proteins in which 2 zones contain amino acid alterations. Two libraries, each library containing alterations within one of the 2 zones, can be prepared.

The variant coat protein fusions will contain one or more alterations including substitutions, additions or deletions relative to the wild type coat protein sequence. Surprisingly, a large number of alterations are possible and are tolerated by the phage while retaining the ability to display polypeptides on the phage surface. Further, the chemical nature of the residue may be changed, i.e. a hydrophobic residue may be altered to a hydrophilic residue or vice versa. Variants containing 2 - 49, preferably 5 - 40, more preferably 7 - 20, altered residues are possible. As demonstrated by the construction of protein P12 below, any of the amino acids of a major coat

protein may be varied, including varying all residues of the coat protein. Fusion proteins containing any mature coat protein sequence or portion thereof which varies from the wild type sequence of the coat protein or portion thereof is within the scope of the invention. Major coat protein variants containing 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 variant residues are possible. Variants containing substitutions or only a few deletions are preferred since these variants will have about the same length as the wild type coat protein sequence. Variants which do not enable surface display of the heterologous polypeptide are selected against during the phage display, panning and selection process.

The construction of such libraries for the M13 major coat protein is described in Example 3; the selection of protein VIII variants which increase the display of hGH or SAV is shown in Example 4. For library construction, the 50-residue protein VIII was divided into five zones encompassing approximately ten contiguous residues each. A library was constructed for each zone of the protein VIII moiety within the hGH-protein VIII fusion encoded by pS349. Most positions were not fully mutated, but variation was allowed at all non-lysine residues. Each library, encoding 10^9 possible protein VIII variants, contained at least 3×10^9 independent transformants.

The libraries were separately cycled through five rounds of binding selection on hGHbp-coated plates. Sequencing of individual clones revealed selectants from libraries encompassing zones 1, 2, and 3 (Fig. 2). The zone 4 and zone 5 sorts yielded contaminants from other libraries. The results suggest that zones 1, 2, and 3 are more tolerant to mutations and thus more suitable for use in the invention, but the results do not exclude the use of mutations in zones 4 and 5 for the purposes of the invention, and mutations in these zones are within the scope of the invention. Repeating the experiment with additional precautions to avoid contamination between libraries will yield variants with mutations in zones 4 and 5 which increase or decrease heterologous protein display.

Selectants were extremely divergent from the wild type sequence, containing seven mutations on average. For zones 1 and 3, a strong consensus was not obtained. Three of four zone 2 selectants were identical, but subsequent clones selected for streptavidin display (see below) yielded little consensus. Only one residue (Ala10) was completely conserved as wild type, six residues (Ala7, Leu14, Ala18, Ile22, Met28, and Val30) showed consensus to wild type. Eight positions showed consensus to a mutant sequence (E2K, D4E, P6F, K8R, FllY, G23R, A27T, and V29Y).

All selectants increased hGH display; Figure 2 shows phage ELISA data for the best selectant from each zone. hGH display with the protein VIII variants produced detectable ELISA signals with phage concentrations in the sub-picomolar range. In contrast, hGH display with wild type protein VIII produced detectable ELISA signals with phage concentrations in the nanomolar

range. Thus, the protein VIII variants increase the signal strength (and decrease the detection limit) by at least three orders of magnitude.

Selection for SAV display was conducted with pooled libraries for zone 1, zone 2, and zone 3 of the protein VIII moiety within the SAV- protein VIII fusion gene. All selectants were from the zone 2 library, but consensus was minimal (Fig. 1B). All selectants increased SAV display. Phage ELISA data for the two best selectants is shown in Fig. 4; the variants provide a 50-fold increase in signal strength when assayed for binding to either anti-SAV antibody or to biotinylated BSA.

In Example 8, hGH mutants with reduced binding affinity for hGHbp were displayed as fusions to either wild type protein VIII or variant protein VIII(1a). Phage ELISA data are shown in Figure 3. As expected, reductions in binding affinity produce corresponding reductions in signal strength. When displayed on wild type protein VIII, the lowest affinity interaction (K_d = 820 nM, 500-fold weaker than wild type) is barely detectable. The same interaction provides an extremely strong signal when displayed on the protein VIII variant. In fact, display of the lowest affinity interaction on the protein VIII variant provides an ELISA signal at least two orders of magnitude greater than that of wild type hGH linked to wild type protein VIII.

Variants with alterations in more than one zone are possible as noted above. Such variants may be obtained by first obtaining a variant with mutations in a single zone (e.g., zone 1) and then using this variant as a template for a second round of selection in which another zone is mutated (e.g., zone 2). Thus, variants obtained by this process will have mutations in two zones (e.g., zones 1 and 2). Alternatively, site-directed mutagenesis can be used to combine mutations from different variants into a single variant. For example, the mutations from a variant with mutations in zone 1 can be introduced into a variant with mutations in zone 2 to produce a new variant with mutations in both zones 1 and 2. This process shown in Example 9.

Obviously, both methods can be extended further to any number of zones which may eventually encompass the entire major coat protein sequence. Thus, it is possible to derive variants of the major coat protein with only minimal homology to the wild type. The potential for extreme variation in sequence in turn provides for extreme variation in function; variants ranging from extremely hyper-functional to extremely hypo-functional can be readily obtained. Thus, the invention can be used to tailor the display level of any heterologous protein fused to a coat protein. It is important to note that while the invention allows for the production of proteins with only limited homology to wild type coat protein, the new proteins are still variants of wild type protein (see definitions above) and are thus within the scope of this invention.

If a variant major coat protein which improves display of a heterologous polypeptide on the surface of phage particles contains multiple mutations relative to wild type, it is also possible to obtain variants which display the heterologous polypeptide at levels intermediate between the levels obtained with the new variant and wild type major coat protein. This can be accomplished by separately back mutating each mutated amino acid of the variant back to the wild type sequence

or to another altered residue. These back mutations will generally reduce display levels of the heterologous polypeptide to levels varying between display levels obtained with the variant and wild type major coat protein. By combining the back mutations, it is possible to tailor display to a desired level which is between that obtained with the variant and wild type major coat protein. This process is shown in Example 10.

By a similar process, it is possible to obtain variants which display at a level below the level of the wild type coat protein. For example, mutations may be made in one or more zones and the libraries produced panned for phage which bind only weakly (weaker than phage displaying wild type fusions). The weaker binding phage will be displaced by phage displaying wild type coat protein fusions and can be isolated and sequenced using known methods.

Mutant coat proteins can be obtained which are hypofunctional for incorporation into the viral coat and thus reduce fusion protein display relative to wild type coat protein. In this case, mutations are made in residues which tend to be conserved as wild type in the above described selections for hyperfunctional variants (e.g. Ala10, Ala7, Leu14, Ala18, Ile22, Met28, and Val30 in protein VIII). Conservation of these residues as wild type during the selection for hyperfunctional variants indicates that mutations at these residues are not well tolerated and will tend to produce hypofunctional variants. Variants obtained through mutations at these sites can then be screened for their ability to display a given fusion protein relative to the wild type coat protein display levels. The heterologous peptide portion of the fusion protein may be any polypeptide or protein which can be captured or bound by a binding partner. Suitable fusions may display an epitope tag or other polypeptide which can be bound by an antibody or binding partner. Hypofunctional variants displaying the fusion at the desired reduced levels relative to wild type can then be used for the construction of libraries of the fusion protein for the purposes of phage display. Although the preferred residues for the production of hypofunctional variants are those which were conserved as wild type, any residue of the coat protein can be mutated and the resulting variant tested for its ability to allow display of a fusion protein. In this way, it is possible to select a display level below that afforded by wild type simply by using the appropriate hypofunctional mutant. As with the hyperfunctional variants described above, several hypofunctional mutations can be combined to produce further reductions in display to levels desired. While, the selection of hypofunctional variants requires a screen rather than a selection, the method is relatively simple since most mutations in proteins cause reductions in activity rather increases and suitable screening procedures are known. Thus, most mutations in the coat protein should be deleterious mutations which result in hypofunctional variants.

The libraries of expression vectors are optionally mixed and used to transform suitable host cells which are subsequently cultured under conditions suitable to form phage or phagemid particles. The phage or phagemid particles are contacted with a target molecule capable of binding to the polypeptide portion of the fusion protein displayed on the surface of the phage, so that at least a portion of the phage particles bind to the target molecule. Particles which contain fusion

proteins displayed in greater numbers on the virus particles will be preferentially bound by the target molecule. Separation of particles which bind to the target from those that do not bind enables one to obtain an enriched library of particles containing fusion proteins which display greater numbers of the heterologous protein on the surface of the particles. This panning process can be repeated multiple times, generally from 2 - 10, preferably 2 - 6, times to obtain further enriched libraries containing clones which contain fusion genes encoding fusion proteins of the heterologous polypeptide linked to the selected variants of the major coat protein, the variants which allow improved incorporation of the fusion protein into the virus coat in the phage display system. This method allows one to select major coat protein variants which are best able to display the particular heterologous polypeptide and optionally other polypeptides as a fusion protein on the surface of the phage particle.

The clones selected by the process described above will have an improved capacity to display the heterologous protein on the surface of the phage in the phage display system. The method is, therefore, generally useful in displaying heterologous polypeptides which are difficult to display using fusion proteins based on wild type coat protein amino acid sequence. The improved fusion protein may be used in monovalent and/or polyvalent phage display systems. In polyvalent systems, the improved fusion protein can be used to increase the number of fusion proteins which are expressed on phage particles or to modulate the number of displayed proteins to a desired range. Phage particles displaying a greater number of fusion proteins will have a greater affinity for target molecules and will be preferentially bound and separated from particles which display fewer heterologous polypeptides as fusion proteins. This is useful in processes for the discovery of weakly binding ligands where the binding affinity of the ligand will be increased by known protein engineering techniques, such as the maturing of antibodies or antibody fragments by humanization, etc. In polyvalent display, the fusion proteins of the invention allow display of only a few to several hundreds of polypeptides on the phage surface. Generally, about 3 to about 50 polypeptides will be displayed in a moderate display system. However, with the fusion protein and phage display system of the invention it is also possible to display about 50, preferably 100 - 900, and up to about 1000 polypeptides or more by selecting a coat protein variant through phage display which is capable of high display numbers.

Vaccination techniques based upon phage expressing antigenic proteins fused to the coat protein have been described (Fanutti, C., et al. (1998) *Biochem. Soc. Trans.*, 26: S8; Jiang, J., et al. (1997) *Infect. Immun.*, 65:4770; Delmastro, P., et al. (1997) *Vaccine*, 15:1276; Galfre, G., et al. (1996) *Methods Enzymol.*, 267:109). This invention can also be used to enhance the effectiveness of phage vaccinations. The variants of the coat protein which increase expression of protein fusions on the surface of phage increase the antigenicity of phage vaccinations. Furthermore, the method can be used to generate variants of the coat protein which stimulate the immune system as haptens. Alternatively, the invention can be used to ameliorate immune response to the phage carrying the antigenic protein.

Alternatively, in monovalent display systems the method and fusion proteins of the invention can be used to precisely tailor display of a protein to a level which is high enough to allow for detection and enrichment of desired affinities, but is low enough to avoid avidity effects associated with polyvalency. Proteins which display polyvalently as fusions to wild type major coat protein can be displayed monovalently as fusions to an appropriate hypo-functional major coat protein variant. Proteins which do not display at all on wild type major coat protein (i.e., cannot be detected as phage-associated entities) can be displayed monovalently as fusions to an appropriate hyper-functional major coat protein variant.

Having obtained a variant major coat protein which improves/tailors display of the heterologous polypeptide on the surface of phage particles, it is then possible to use conventional phage display technologies to construct libraries of variants of the originally displayed heterologous polypeptide and select for a desired property, e.g., binding, enzymatic activity, etc. The fusion protein of the invention, containing a selected variant of the major coat protein of the phage which provides the desired display characteristics, can be used to replace a fusion protein in conventional phage display systems where the conventional fusion protein contains the wild type amino acid sequence of the major coat protein or coat protein fragment. Replacement of the conventional fusion protein with the variant fusion protein of the invention improves the display of heterologous polypeptide in the phage display system. That is, the coat protein portion has been optimized for the polypeptide which is displayed as a fusion protein.

Further, it is possible to replace the original heterologous polypeptide in the new fusion protein obtained as described above with a second different heterologous polypeptide and maintain the benefits of improved incorporation of the fusion protein into virus particles. See Example 7. In this aspect of the invention, the fusion gene which encodes the original polypeptide/major coat protein variant fusion obtained by phage display panning and selection as discussed above is modified to replace the gene encoding the first polypeptide with a gene encoding a second polypeptide of interest. Conventional phage display libraries can then be constructed in which one or more residues of the second polypeptide are varied and selected by phage display panning and selection to obtain variants of the second polypeptide with the desired (e.g., improved) binding properties. This result is also surprising since the variant coat protein portion of the fusion protein was originally selected for the ability to display a different (e.g., the original) polypeptide. Nevertheless, it has been discovered that a fusion protein containing a variant coat protein portion which has been selected for improved display of a heterologous polypeptide will also generally provide improved display of other unrelated heterologous polypeptides as well, even other polypeptides containing multiple subunits. The fusion protein containing the variant coat protein portion can, therefore, be used in general phage display systems.

The phage display system of the invention can also be used to isolate polypeptides which are produced as therapeutic polypeptides using conventional recombinant DNA technology. In this embodiment, the method described above is used to identify a fusion protein containing a desired

coat protein variant portion for use in phage display. The heterologous polypeptide portion of the fusion protein may be the desired product polypeptide itself or may be a different polypeptide where this phage display step is used to select for the coat protein variant portion providing improved surface display as described above. Using a gene for the selected coat protein variant portion in the fusion gene together with a gene for a heterologous polypeptide, one can then use phage display to optimize and select from a library of different potential product heterologous polypeptide sequences to obtain a product polypeptide sequence. This product polypeptide sequence is then cloned into an expression plasmid containing a transcription regulatory element operably linked to a gene fusion encoding the product polypeptide. Expression of the gene fusion in mammalian or bacterial cells yields the product polypeptide using well known recombinant technology. See Sambrook *et al.*

It is also within the scope of the invention to prepare fusion proteins of a heterologous polypeptide and a portion of a phage coat protein, which is not necessarily the major coat protein of the phage, used to display the polypeptide. In this embodiment, for example, a minor coat protein such as coat protein III of a filamentous phage is mutated to form families and libraries of fusion proteins and phage variants as described above and phage display selection and panning are used to obtain specific phage displaying a fusion protein of a heterologous polypeptide and at least a portion of a coat protein variant. Preferably, the coat protein portion is a mutant having at least one altered residue in the transmembrane domain or in the cytoplasmic domain of the coat protein. With respect to coat protein III, these altered residues will preferably be in the region of residues 377 to 406 as counted from the amino terminal end of the mature coat protein III (Marvin, D. A., Filamentous phage structure, infection, and assembly, *Current Opinion in Structural Biology*, 1998, 8:150-158). Coat protein III variants may contain a plurality of variant residues as generally described above for major coat proteins.

Suitable gene III vectors for display of polypeptides include fUSE5 (Scott, J. K. and Smith G. P. (1990). Searching for peptide ligands with an epitope library. *Science* 249, 386-390); fAFF1 (Cwirla., S. E., Peters, E. A., Barrett, R. W., and Dower, W. J. (1990). Peptides of phage: A vast display library of peptides for identifying ligands. *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382); fd-CAT1 (McCafferty, J., Griffiths, A., D., Winter, G., and Chiswell, D., J. (1990). Phage antibodies: Filamentous phage displaying antibody variable domains. *Nature* (*London*) 348, 552-554); m663 (Fowlkes, D., Adams, M., Fowler, V., and Kay, B. (1992). Mutlipurpose vectors for peptide expression on the M13 viral surface. *Biotechniques* 13, 422-427); fdtetDOG, pHEN1 (Hoogenboom, H., Griffiths, A., Johnson, K., Chisswell, D., Hudson, P., and Winter, G. (1991). Multi-subunit proteins on the surfaces of filamentous phage: Methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res.* 19:4133-4137); pComb3 (Gram, H., Marconi, L. A., Barbas, C. F., Collet, T. A., Lerner, R. A., and Kang, A. S. (1992) *In vitro* selection and affinity maturation of antibodies from a naive combinatorial immunoglobulin library. *Proc. Natl. Acad. Sci. U.S.A.* 89, 3576-3580); pCANTAB 5E (Pharmacia); and LamdaSurfZap

(Hogrefe, H. H., Amberg, J. R., Hay, B. N., Sorge, J. A., and Shopes, B. (1993) Cloning in a bacteriophage lambda vector for the display of binding of binding proteins on filamentous phage. *Gene* 137, 85-91).

Phage display methods for proteins, peptides and mutated variants thereof, including constructing a family of variant replicable vectors containing a transcription regulatory element operably linked to a gene fusion encoding a fusion polypeptide, transforming suitable host cells, culturing the transformed cells to form phage particles which display the fusion polypeptide on the surface of the phage particle, contacting the recombinant phage particles with a target molecule so that at least a portion of the particle bind to the target, separating the particles which bind from those that do not bind, are known and may be used with the method of the invention. See U.S. 5,750,373; WO 97/09446; U.S. 5,514,548; U.S. 5,498,538; U.S. 5,516,637; U.S. 5,432,018; WO 96/22393; U.S. 5,658,727; U.S. 5,627,024; WO 97/29185; O'Boyle et al., 1997, *Virology*, 236:338-347; Soumillion et al., 1994, *Appl. Biochem. Biotech.*, 47:175-190; O'Neil and Hoess, 1995, *Curr. Opin. Struct. Biol.*, 5:443-449; Makowski, 1993, *Gene*, 128:5-11; Dunn, 1996, *Curr. Opin. Struct. Biol.*, 7:547-553; Choo and Klug, 1995, *Curr. Opin. Struct. Biol.*, 6:431-436; Bradbury and Cattaneo, 1995, *TINS*, 18:242-249; Cortese et al., 1995, *Curr. Opin. Struct. Biol.*, 6:73-80; Allen et al., 1995, *TIBS*, 20:509-516; Lindquist and Naderi, 1995, *FEMS Micro. Rev.*, 17:33-39; Clarkson and Wells, 1994, *Tibtech*, 12:173-184; Barbas, 1993, *Curr. Opin. Biol.*, 4:526-530; McGregor, 1996, *Mol. Biotech.*, 6:155-162; Cortese et al., 1996, *Curr. Opin. Biol.*, 7:616-621; McLafferty et al., 1993, *Gene*, 128:29-36.

The heterologous polypeptide portion of the fusion protein may contain as few as 4-10 or up to 20-30 amino acid residues and even up to about 50 –80 residues. These smaller peptides are useful in determining the antigenic properties of the peptides, in mapping the antigenic sites of proteins, etc. The heterologous polypeptide may also contain one or more subunits containing at least about 100 amino acid residues which may be folded to form a plurality of rigid secondary structures displaying a plurality of amino acids capable of interacting with the target. If the heterologous polypeptide portion of the fusion protein is mutated to form a library and subjected to phage display selection, it is preferred that polypeptide be mutated at codons corresponding to the amino acids capable of interacting with the target so that the integrity of the rigid secondary structures will be preserved. The residues can be determined by alanine scanning mutagenesis, for example. U.S. 5,580,723 and U.S. 5,766,854.

The heterologous polypeptide portion may also be a protein, preferably a mammalian protein, such as a cytokine, and the protein may be selected from human growth hormone(hGH), N-methionyl human growth hormone, bovine growth hormone, parathyroid hormone, thyroxine, insulin A-chain, insulin B-chain, proinsulin, relaxin A-chain, relaxin B-chain, prorelaxin, glycoprotein hormones such as follicle stimulating hormone(FSH), thyroid stimulating hormone(TSH), leutinizing hormone(LH), glycoprotein hormone receptors, calcitonin, glucagon, factor VIII, an antibody, lung surfactant, urokinase, streptokinase, human tissue-type plasminogen

activator (t-PA), bombesin, coagulation cascade factors including factor VII, factor IX, and factor X, thrombin, hemopoietic growth factor, tumor necrosis factor-alpha and -beta, enkephalinase, human serum albumin, mullerian-inhibiting substance, mouse gonadotropin-associated peptide, a microbial protein, such as betalactamase, tissue factor protein, inhibin, activin, vascular endothelial growth factor (VEGF), receptors for hormones or growth factors; integrin, thrombopoietin (TPO), protein A or D, rheumatoid factors, nerve growth factors such as NGF- alpha, platelet-growth factor, transforming growth factors (TGF) such as TGF-alpha and TGF-beta, insulin-like growth factor-I and -II, insulin-like growth factor binding proteins, CD-4, DNase, latency associated peptide, erythropoietin (EPO), osteoinductive factors, interferons such as interferon-alpha, ,-alphacon-1, -beta, and -gamma, colony stimulating factors (CSFs) such as M-CSF, GM-CSF, and G-CSF, interleukins (ILs) such as IL-1, IL-2, IL-3, IL-4, IL-6, IL-8, IL-10, IL-12, superoxide dismutase; decay accelerating factor, viral antigen, HIV envelope proteins such as GP120, GP140, atrial natriuretic peptides A, B or C, Apo2L, novel erythropoiesis stimulating protein (NESP), ancestim, keratinocyte growth factor (KGF), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin, IL-1 receptor antagonist (IL-1ra), soluble tumor necrosis factor-a receptor type I (sTNF-RI), immunoglobulins, as well as variants and fragments of any of the above-listed proteins

The heterologous polypeptide portion may also include a molecular tag, also known as an epitope tag, for identifying and/or capturing and purifying the fusion protein. For example, the tag may be Herpes simplex virus glycoprotein D (Paborsky et al. 1990, *Protein Engineering*, 3:547-553) which can be used to affinity purify the fusion protein through binding to an anti-gD antibody, protein A or a fragment thereof (Li et al. (1998) Mol Biotech., 9:187), a polyhistidine tag, e.g., $(his)_6$ (Sporeno et al., 1994, *J. Biol. Chem.*, 269:10991-10995; Stuber et al., 1990, *Immunol. Methods*, 4:121-152, Waeber et al., 1993, *FEBS Letters*, 324:109-112), etc., which can be used to identify and/or purify the fusion protein through binding to a metal ion (Ni) column (QIAEXPRESS Ni-NTA protein Purification System, Quiagen, Inc.). Other affinity tags known in the art may be used.

In a particularly preferred embodiment, the heterologous polypeptide portion of the fusion protein is the light chain or the heavy chain of an antibody or fragments thereof, such Fab, $F(ab')_2$, Fv, diabodies, linear antibodies, etc. The polypeptide may also be a single chain antibody (scFv). The preparation of libraries of antibodies or fragments thereof is well known in the art and any of the known methods may be used to construct a family of transformation vectors which may be transformed into host cells using the method and fusion protein of the invention. Libraries of antibody light and heavy chains in phage (Huse et al., 1989, *Science*, 246:1275) and as fusion proteins in phage or phagemid are well known and can be prepared according to known procedures. See Vaughan et al., Barbas et al., Marks et al., Hoogenboom et al., Griffiths et al., de Kruif et al., noted above, and WO 98/05344; WO 98/15833; WO 97/47314; WO 97/44491; WO 97/35196; WO 95/34648; U.S. 5,712.089; U.S. 5,702,892; U.S. 5,427,908; U.S. 5,403,484; U.S.

5,432,018; U.S. 5,270,170; WO 92/06176; U.S. 5,702,892. Reviews have also published. Hoogenboom, 1997, *Tibtech*, 15:62-70 ; Neri et al., 1995, *Cell Biophysics*, 27:47; Winter et al., 1994, *Annu. Rev. Immunol.*, 12:433-455; Soderlind et al., 1992, *Immunol. Rev.*, 130:109-124; Jefferies, 1998, *Parasitology*, 14:202-206.

Specific antibodies contemplated as the heterologous polypeptide portion include antibodies which bind to human leukocyte surface markers, cytokines and cytokine receptors, enzymes, etc. Specific leukocyte surface markers include CD1a-c, CD2, CD2R, CD3-CD10, CD11a-c, CDw12, CD13, CD14, CD15, CD15s, CD16, CD16b, CDw17, CD18-C41, CD42a-d, CD43, CD44, CD44R, CD45, CD45A, CD45B, CD45O, CD46-CD48, CD49a-f, CD50-CD51, CD52, CD53-CD59, CDw60, CD61, CD62E, CD62L, CD62P, CD63, CD64, CDw65, CD66a-e, CD68-CD74, CDw75, CDw76, CD77, CDw78, CD79a-b, CD80-CD83, CDw84, CD85-CD89, CDw90, CD91, CDw92, CD93-CD98, CD99, CD99R, CD100, CDw101, CD102-CD106, CD107a-b, CDw108, CDw109, CD115, CDw116, CD117, CD119, CD120a-b, CD121a-b, CD122, CDw124, CD126-CD129, and CD130.

Other antibody binding targets include cytokines and cytokine superfamily receptors, hematopoietic growth factor superfamily receptors and preferably the extracellular domains thereof, which are a group of closely related glycoprotein cell surface receptors that share considerable homology including frequently a WSXWS domain and are generally classified as members of the cytokine receptor superfamily (see e.g. Nicola *et al.*, *Cell*, 67:1-4 (1991) and Skoda, R.C. *et al. EMBO J.* 12:2645-2653 (1993)). Generally, these targets are receptors for interleukins (IL) or colony-stimulating factors (CSF). Members of the superfamily include, but are not limited to, receptors for: IL-2 (b and g chains) (Hatakeyama *et al.*, *Science*, 244:551-556 (1989); Takeshita *et al.*, *Science*, 257:379-382 (1991)), IL-3 (Itoh *et al.*, *Science*, 247:324-328 (1990); Gorman *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:5459-5463 (1990); Kitamura *et al.*, *Cell*, 66:1165-1174 (1991a); Kitamura *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:5082-5086 (1991b)), IL-4 (Mosley *et al.*, *Cell*, 59:335-348 (1989), IL-5 (Takaki *et al.*, *EMBO J.*, 9:4367-4374 (1990); Tavernier *et al.*, *Cell*, 66:1175-1184 (1991)), IL-6 (Yamasaki *et al.*, *Science*, 241:825-828 (1988); Hibi *et al.*, *Cell*, 63:1149-1157 (1990)), IL-7 (Goodwin *et al.*, *Cell*, 60:941-951 (1990)), IL-9 (Renault *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:5690-5694 (1992)), granulocyte-macrophage colony-stimulating factor (GM-CSF) (Gearing *et al.*, *EMBO J.*, 8:3667-3676 (1991); Hayashida *et al.*, *Proc. Natl. Acad. Sci. USA*, 244:9655-9659 (1990)), granulocyte colony-stimulating factor (G-CSF) (Fukunaga *et al.*, *Cell*, 61:341-350 (1990a); Fukunaga *et al.*, *Proc. Natl. Acad. Sci. USA*, 87:8702-8706 (1990b); Larsen *et al.*, *J. Exp. Med.*, 172:1559-1570 (1990)), EPO (D'Andrea *et al.*, *Cell*, 57:277-285 (1989); Jones *et al.*, *Blood*, 76:31-35 (1990)), Leukemia inhibitory factor (LIF) (Gearing *et al.*, *EMBO J.*, 10:2839-2848 (1991)), oncostatin M (OSM) (Rose *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:8641-8645 (1991)) and also receptors for prolactin (Boutin *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:7744-7748 (1988); Edery *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:2112-2116

(1989)), growth hormone (GH) (Leung *et al., Nature,* 330:537-543 (1987)), ciliary neurotrophic factor (CNTF) (Davis *et al., Science*, 253:59-63 (1991) and c-Mpl (M. Souyri *et al., Cell* 63:1137 (1990); I. Vigon *et al., Proc. Natl. Acad. Sci.* 89:5640 (1992)). Still other targets for antibodies made by the invention are erb2, erb3, erb4, IL-10, IL-12, IL-13, IL-15, tumor necrosis factor alpha, thrombin, etc. The variant coat protein fusions and the variant heterologous polypeptides and libraries containing the same can be prepared using conventional mutagenesis techniques. These methods include but are not limited to oligonucleotide-mediated mutagenesis and cassette mutagenesis.

The heterologous polypeptide may be linked to the coat protein or portion thereof through a peptide linker. A linker peptide segment will generally vary in length from about 3 to about 50 amino acid residues, preferably from 5 to 30 residues, more preferably from 10 to 25 residues. Further, the net charge on the linker segment is preferably positive. The identity of and order of the amino acid residues is optional, although one or more specific sequences of the linker peptide segment will generally provide better display of the heterologous polypeptide. The method of this invention can also be used to modulate display levels of a fusion protein by mutating the linker between the fused protein and the coat protein and selecting linkers which afford the desired level of display. In this embodiment, a library of linker segment variants is made by mutating a linker sequence template and the linker sequences which give the best display on phage are selected using phage display selection, for example, an affinity selection for binding to the displayed heterologous polypeptide. Linkers which allow for greater numbers of displayed polypeptides will be selected based on increased affinity for the affinity matrix.

To date, researchers have used specific linkers designed to provide desired attributes. Linkers have been designed to provide flexibility (Wung et al. (1997) *J. Immunol. Methods* 204:33-41), such as the Gly-Ala_3 (Holmes et al. (1996) *Protein Pept. Lett.* 3:415-422) or Gly_4Ser_3 linkers (Micheal et al. (1996) *Immunotechnology* 2:47-57) and to incorporate sites for specific proteolysis (Lucic, et al. (1998) *Australia. J. Biotechnol.* 61:95-108; Matthews, D.J. and Wells, J.A. (1993) *Science*, 26:1113-1117). Considerations for linker optimization include, among other factors, resistance to proteolysis, distance from the phage particle to the fused protein, and conformational effects of the linker upon fusion protein activity. The large number of variables involved makes the selection method of the invention an attractive approach. For example, selection of a linker for increased display of hGH on protein VIII (Fig. 7A) results in increased display relative to a designed Gly/Ser linker (Fig. 8A). A similar selection was also used to select linkers (Fig. 7B) for SAV display (Fig. 8B). In the SAV display, the selection was performed with SAV fused to a protein VIII variant (protein VIII(1e)) which had been previously selected for increased display of SAV (Example 4). This example demonstrates that optimized linkers can be combined with optimized coat protein variants, for example protein VIII variants, to obtain a desired display level, in a manner similar to the combination of different protein VIII variants described above.

The methods described above with reference to obtaining a gene encoding a heterologous polypeptide, variant polypeptide or a fusion protein containing at least a portion of a phage coat protein and a heterologous peptide or variant and isolating the same, oligonucleotide-mediated and cassette mutagenesis, cleaving DNA using restriction enzymes, ligation, separation and selection of DNA using electrophoresis, purification and transformation (e.g., electroporation) procedures, library construction, suitable host or recipient cells and cell concentrations, etc. and the prior art methods noted above may be used in this embodiment of the invention and the description thereof is incorporated here specifically with respect to this embodiment.

C. Carboxyl-terminal display and more new phage coat proteins and fusions thereof

Another aspect of the invention is the carboxyl-terminal (C-terminal) display of a heterologous polypeptide on the surface of a filamentous phage using protein fusions with protein III or protein VIII. C-terminal display has been reported on protein VI of M13 (Jespers, L. et al., 1995, Biotechnology 13:378-382). This paper states that protein VI is distinct from proteins III and VIII in its ability to allow for the attachment of polypeptides at the C-terminus. Surprisingly, as a part of this invention, it has been discovered that C-terminal display is also possible with fusions to protein III and VIII. The invention, therefore, allows the C-terminal display of a heterologous polypeptide or library of polypeptides in a manner similar to display at the N-terminus (N-terminal display) of a phage coat protein. In this aspect of the invention, the C-terminal display may be accomplished using a wild type protein III/VIII or a mutant protein III/VIII as described above where the phage display process was applied to the coat protein sequence itself.

Any of the methods of phage or phagemid display, creating coat protein variants and protein fusions thereof with a heterologous polypeptide, libraries of such variants and fusion proteins, expression vectors encoding the variants and protein fusions, libraries of the vectors, a library of host cells containing the vectors, methods for preparing and panning the same to obtain binding polypeptides, etc. described above with reference to N-terminal display may also be used in this aspect of the invention for C-terminal display and the descriptions above are hereby incorporated here and should be considered as part of the description of C-terminal display of the invention.

The invention allows one to evolve new virus particles containing non-wild type coat proteins and coat protein fusions.

The variant protein III/VIII fusion proteins will contain one or more alterations including substitutions, additions or deletions relative to the wild type coat protein sequence. Again, it is surprisingly, that a large number of alterations are possible and are tolerated by the phage while retaining the ability to display polypeptides on the phage surface, in this case as C-terminal fusions. The chemical nature of the residue may be changed, i.e. a hydrophobic residue may be altered to a hydrophilic residue or vice versa. Variants containing 2 - 50, preferably 5 - 40, more preferably 7 -

20, altered residues are possible. Fusion proteins containing any mature coat protein sequence or portion thereof which varies from the wild type sequence of the coat protein or portion thereof is within the scope of the invention. Major coat protein variants containing 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49 or 50 variant residues are possible. This aspect of the invention allows one to design a coat protein which is any coat protein other that the wild type coat protein and select for C-terminal fusion proteins which display on the surface of phage. Variants containing substitutions are preferred since these variants will have about the same length as the wild type coat protein sequence. However, deletions of residues to prepare shorter coat proteins and protein fusions thereof are within the scope of the invention. Preferably, the first few residues will be deleted, more preferably N-terminal or C-terminal residues 1 to about 5 can be deleted. Variants which do not enable surface display of the heterologous polypeptide are selected against during the phage display, panning and selection process.

As with N-terminal display described above, it is also possible to produce libraries in which amino acids residues within desired segments of the coat protein are varied to obtain a library of coat protein variants having amino acid additions, substitutions or deletions within defined regions of the coat protein. As an example, the coat protein may be divided into an arbitrary number of zones, generally 2-10 zones, and a library constructed of variants within one or more of the zones. The mature coat proteins of M13, f1 and fd phage, for example, contain 50 amino acids and might be divided into 10 zones of 5 amino acid residues each or into zones with unequal numbers of residues in each zone, e.g. zones containing 15, 10, 9, and 8 residues. Zones corresponding to the cytoplasmic, transmembrane and periplasmic regions of the coat protein may be used. A separate library may be constructed for each of the zones in which amino acid alterations are desired. If fusion proteins are desired in which the coat protein variant has an amino acid alteration in zone 1, for example, a single library may be constructed in which one or more of the amino acid residues within zone 1 is varied. Alternatively, one may wish to produce fusion proteins in which 2 zones contain amino acid alterations. Two libraries, each library containing alterations within one of the 2 zones, can be prepared.

Preferably, the heterologous polypeptide is attached to the coat protein or variant thereof through a linker peptide. The linker may contain any number of residues which allow C-terminal display, and will generally contain about 4 to about 30, preferably about 8 to about 20, amino acid residues. The linker may contain any of the naturally occurring residues, although linkers containing predominantly (greater than 50%) glycine and/or serine are preferred. The optimum linker composition and length for display of a particular polypeptide may be selected using phage display as described above and demonstrated in the examples. For example, phage libraries each containing a different linker length may be constructed and phage selection and panning used to isolate the amino acid composition of the linker of any length which optimizes expression and display of the heterologous polypeptide.

As with N-terminal display described above, if a variant coat protein which improves display of a heterologous polypeptide on the surface of phage particles contains multiple mutations relative to wild type, it is also possible to obtain variants which display the heterologous polypeptide at levels intermediate between the levels obtained with the new variant and wild type coat protein. This can be accomplished by separately back mutating each mutated amino acid of the variant back to the wild type sequence or to another altered residue. These back mutations will generally reduce display levels of the heterologous polypeptide to levels varying between display levels obtained with the variant and wild type coat protein. By combining the back mutations, it is possible to tailor display to a desired level which is between that obtained with the variant and wild type coat protein.

By a similar process, it is possible to obtain variants which display at a level below the level of the wild type coat protein. For example, mutations may be made in one or more zones and the libraries produced panned for phage which bind only weakly (weaker than phage displaying wild type fusions). The weaker binding phage will be displaced by phage displaying wild type coat protein fusions and can be isolated and sequenced using known methods.

Mutant coat proteins can also be obtained which are hypofunctional for incorporation into the viral coat and thus reduce fusion protein display relative to wild type coat protein. In this case, mutations are made in residues which tend to be conserved as wild type in the above described selections for hyperfunctional variants. Conservation of these residues as wild type during the selection for hyperfunctional variants indicates that mutations at these residues are not well tolerated and will tend to produce hypofunctional variants. Variants obtained through mutations at these sites can then be screened for their ability to display a given fusion protein relative to the wild type coat protein display levels. Hypofunctional variants displaying the fusion at the desired reduced levels relative to wild type can then be used for the construction of libraries of the fusion protein for the purposes of phage display. Although the preferred residues for the production of hypofunctional variants are those which were conserved as wild type, any residue of the coat protein can be mutated and the resulting variant tested for its ability to allow display of a fusion protein. In this way, it is possible to select a display level below that afforded by wild type simply by using the appropriate hypofunctional mutant. As with the hyperfunctional variants described above, several hypofunctional mutations can be combined to produce further reductions in display to levels desired. While, the selection of hypofunctional variants requires a screen rather than a selection, the method is relatively simple since most mutations in proteins cause reductions in activity rather increases and suitable screening procedures are known. Thus, most mutations in the coat protein should be deleterious mutations which result in hypofunctional variants.

C-terminal display is useful to display cDNA libraries on the surface of phage particles. mRNA can be purified from a tissue source of choice and double stranded cDNAs synthesized using standard techniques (Sambrook et al., 1989, Molecular Cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor, New York). A phagemid or phage vector (or a plurality thereof)

containing an open reading frame is then constructed using well established techniques disclosed in Sambrook et al. and the phage and phagemid display references described above and the cDNAs are ligated into the vectors at the 3' end of the coat protein gene. Host cells are then transformed, preferably by electroporation, with the library of vectors and phage particles displaying heterologous polypeptides corresponding to the cDNA library members are obtained (with superinfection of helper phage for phagemid vectors). The C-terminal phage display library obtained may be panned and analyzed using conventional phage display techniques.

The C-terminal display of the invention is also useful to display intracellular, preferably mammalian intracellular, proteins or fragments thereof and polypeptides which are difficult to display using N-terminal display. C-terminal display is, therefore, a complementary display technique to N-terminal display. Intracellular proteins may be difficult to display in a correctly folded form using N-terminal display due to the difference in redox environment in which intracellular proteins normally exist relative to the environment in which secreted proteins fold and form disulfide bonds. The cytoplasm is a reducing environment whereas the periplasm is an oxidizing environment. C-terminal heterologous fusion proteins migrate to the periplasm as in normal phage particle assembly. However, since the heterologous polypeptide remains on the intracellular side of the periplasmic membrane, an intracellular polypeptide may correctly fold prior to incorporation into a phage particle. During assembly of the phage or phagemid particle, the C-terminal fusion protein is incorporated into the particle and displays the heterologous polypeptide on the surface thereof.

C-terminal display bypasses secretion problems encountered with N-terminal display systems. With N-terminal display, it is generally thought that the heterologous polypeptide on the N-terminus must pass through a pore-like structure in the periplasmic membrane in order to enter the periplasmic space with the C-terminus remaining as an anchor in the membrane. The fusion protein is then assembled into a phage particle from the membrane. Using C-terminal display, it is not necessary to have the fusion protein secrete into the host cell periplasm in order to assemble phage particles. C-terminal display is, therefore, useful to display any heterologous polypeptide and is particularly useful to display polypeptides which are difficult to display using N-terminal phage display techniques.

C-terminal display can also be used to display a foreign polypeptide which is secreted into the periplasm during virus particle assembly. For example, by constructing a library of potential membrane proteins and selecting members of the library capable of functioning as coat proteins, it is possible to apply selective pressure to the library and evolve coat proteins which have a foreign polypeptide as a C-terminal fusion and which orient in the cell membrane with the foreign polypeptide in the periplasm and the N-terminus of the fusion protein in the cytoplasm. Such fusion proteins preferably have a positively charged N-terminal portion as a cytoplasmic region and a hydrophobic core portion as a transmembrane region. Such a structure resembles a bacterial secretion signal. Some library members will function as secretion signals and insert into the

bacterial membrane with the N-terminus in the cytoplasm and the C-terminus in the periplasm (inverse coat proteins). Some of the fusions which can insert into the membrane will incorporate into assembling virus particles by virtue of favorable interactions with the phage or phagemid coat. Suitable libraries may be designed in multiple stages. For example, inverse coat proteins may be selected from a library (or libraries) of potential coat proteins by using an epitope tag fused to the C-terminus of the library of coat proteins. After phage particle assembly, antibody binding to the epitope tag is used to isolate members of the virus particle library which display the tag on the surface of the particle. The particles which bind can be isolated/selected and cloned using conventional phage display techniques. In a second step, one of the selectants can be further evolved by phage display to select for improved incorporation into a particle coat. Again, one or more libraries can be constructed to vary different regions of the coat protein to select for those proteins which best display the epitope tag or other protein on the surface of the particle. Coat protein fusions and virus particles prepared according to the invention provide a diversity of tools which are useful to evaluate virus structure and assembly processes, to map the antibody binding epitopes on proteins, to affinity mature binding proteins such as antibodies and fragments thereof to provide binding proteins with higher binding affinities, to produce polypeptides which bind to active and allosteric sites on enzymes, etc. including all of the uses for which phage display is currently employed.

All patent and literature references cited above are incorporated herein by reference in their entirety.

EXAMPLES

Example 1 - Construction of *E. coli* SS320

The new cell line SS320 was prepared by bacterial mating in which the F' episome was transferred from XL1-BLUE cells to MC1061cells according to known protocols (J. H. Miller, 1972, *Experiments in Molecular Biology*, p190). More specifically, the SS320 cells can be obtained using the following steps:

-Grow 1.0 mL cultures of MC1061 and XL1-BLUE in LB broth to OD600 = 0.5 (single colonies from freshly streaked plates. MC1061 was streaked on LB plates. XL1-BLUE was streaked on LB/tetracycline (10μg/mL)).

-Mix 0.5 mL of each culture and grow 1 hour at 37°C with slow shaking (50 rpm on a rotary shaker). After mating for 1 hour, agitate at 250 rpm to disrupt the mating.

-Plate dilutions on LB/tetracycline(10μg/mL)/streptomycin(10μg/mL).

-MC1061 carries a streptomycin resistance chromosomal marker while the F' plasmid of XL1-BLUE confers tetracycline resistance. Thus, only the mating progeny (MC1061 harboring the XL1-BLUE F' episome will be resistant to both antibiotics. The unmixed cultures of MC1061

and XL1-BLUE can be plated on the selective media as controls, since neither parent is doubly resistant.

-The resulting strain (SS320) can be used for electroporation and phage production.

The genotypes of the starting MC1061 cells (available from Bio-Rad Laboratories, Inc.) and XL1-BLUE episome (available from Stratagene, Inc. in XL1-BLUE cells) and the resulting SS320 cells are as follows:

XL1-BLUE F' episome

F'::Tn10 *proA*$^+$*B*$^+$*lacI*^{q}*D(lacZ)M15*

MC1061

F-*araD139D(ara-leu)7696galE15galK16D(lac)X74rpsL*(Str^r)*hsdR2*($r_K^-m_K^+$)*mcrAmcrB1*

SS320

F'::Tn10 *proA*$^+$*B*$^+$*lacI*^{q}*D(lacZ)M15/*

F-*araD139D(ara-leu)7696galE15galK16D(lac)X74rpsL*(Str^r)*hsdR2*($r_K^-m_K^+$)*mcrAmcrB1*

Various *E. coli* strains were evaluated for cell survival and viability after standard washing steps were performed to prepare the cells for electroporation. *E. coli* were grown in 250 mL cultures and prepared for electroporation as previously described. The total number of colony forming units were titered before and after the wash procedure and the results are shown below.

Strain	Before wash (cfu)	After wash (cfu)	Survival (After/Before)
SS320	4.8×10^{12}	4.6×10^{12}	0.96
TB1	1.7×10^{11}	1.9×10^{11}	1.1
JM101	4.5×10^{11}	3.6×10^{10}	0.08
JM107	2.3×10^{11}	1.0×10^{10}	0.045
JM109	1.5×10^{11}	1.2×10^{10}	0.08

The transformation yield achieved with different *E. coli* strains at a fixed concentration of DNA is dependent on the concentration of viable *E. coli* cells in the electroporation reaction. It has now been discovered that the maximum concentration of viable cells which can be achieved with a given strain is dependent upon the resistance of that strain to the washing steps involved in the preparation of electrocompetent cells. The suitability of a particular bacterial (e.g., *E. coli)* strain for electroporation can be easily determined using the following procedure.

Grow a 250 mL culture of bacteria to OD600 = 0.6 in a 1-L baffled flask. Remove a small aliquot and plate dilutions on appropriate media to determine the total number of viable cells; this

number is the input of cells (input, I). Follow the standard procedures for production of electrocompetent cells as described in Example 2 (scaled down appropriately for the volume of culture used). Plate dilutions of the final preparation of electrocompetent cells on appropriate media to determine the total number of viable cells; this is the number of cells surviving the electrocompetent cell preparation procedure (survivors, S). Divide S by I to determine the ratio of survivors to input (S/I). For a strain ideally suited for electroporation (i.e. a strain which gives the highest transformation yield in comparison with other strains at a fixed DNA concentration), the ratioS/I should be equal to one. This indicates that all the input cells survived the electrocompetent cell preparation procedure. A reduction in the ratio S/I corresponds to a reduction in the concentration of viable cells in the electrocompetent cell preparation; this in turn results in a reduction in the transformation yield. Thus, for the highest transformation yields at a given DNA concentration, a strain with the highest S/I value should be used.

Example 2 – Preparation of *E. coli* for electroporation

Electroporation competent cells were prepared as described below:

1. Inoculate 1 mL 2xYT media(5 mg/mL tetracycline) with SS-320 from a fresh LB/tet plate. Grow about 6 hours and inoculate 50 mL 2xYT/tetracyline in a 500-mL flask; grow overnight.

2. Inoculate 6 x 900 mL Superbroth (5 mg/mL tetracycline) in 2-L baffled flasks with 5 mL from above culture and grow cells to OD_{600} = 0.6-0.8 at 37°C, 200 rpm.

3. Chill three flask on ice (shake periodically). Further steps were performed in a cold room, on ice, with prechilled solutions and equipment.

4. Centrifuge 5.5K/5 min in a SORVALL GS3 ROTOR and decant all supernatant. Add culture from remaining three flasks to same tubes; respin and decant.

5. Resuspend in equal volume of 1 mM HEPES, pH7.4 by swirling or stirring. Centrifuge 5.5K/10 min and decant supernatant.

6. Resuspend in equal volume of 1 mM HEPES as in (5) above. Centrifuge 5.5K/10 min and decant supernatant. Resuspend each pellet in 100 mL of 10% (v/v) glycerol (filter sterilized; ultrapure glycerol (Gibco BRL #15514-011)).

7. Centrifuge 5.5K/15 min and decant all supernatant. Resuspend in minimum volume of 10% glycerol. Using about 3 mL of 10% glycerol for 5 L of starting culture produces about 12 mL of concentrated cells with about 3-4 x 10^{11} cfu/mL.

Example 3 – Mutagenesis Fill-in

The mutagenesis reaction was conducted using the procedure described in U.S. 5,750,373 with the changes shown below:

1) Kinase oligo

4 µL oligo (330 ng/mL stock; *i.e.*, $A_{260} = 10$)

4 µL 10X TM buffer (0.5M tris pH7.5, 0.1M $MgCl_2$)

4 µL 10mM ATP

2 µL 100mM DTT

24 µL H_2O

2 µL kinase (NEB, 10 U/µL)

40 µL

-incubate at 37°C for 0.5 hour.

2) Anneal oligo/template

40 µg kunkel template

1.2 µg kinased oligo (*i.e.* 40 µL from above kinase reaction; oligo/template = 3)

25 µL 10X TM buffer

add H2O to 250 µL final volume

-incubate at 90°C for 2 minutes, 50°C for 3 minutes.

3) Fill-in

-add:

1 µL 100mM ATP

10 µL 25mM dNTPs (25mM each dATP, dCTP, dGTP, dTTP)

15 µL 100mM DTT

6 µL T4 ligase (NEB, 400 U/µL)

3 µL T7 polymerase (NEB, 10 U/µL)

-incubate at 20°C for 3 hours.

Example 4 – *E. coli* Electroporation

Electroporation was conducted as described below:

1. Extract fill-in reaction with an equal volume of phenol/$CHCl_3$. Extract with an equal volume of $CHCl_3$. Purify and desalt DNA (60 micrograms) using QIAquick gel extraction kit (QIAGEN). Use two columns for each reaction. Follow wash and elution procedures as outlined by QIAGEN; elute each column with 30 µL of H_20 to provide a final theoretical yield of 80 µg fill-in product (40 µg single strand DNA converted to double strand) in 60 µL of H_2O.

2. Electroporate DNA (60 µL) into 350 µL of competent *E. coli* SS320. Use 0.2 cm gaps cells with following settings: 2.5 kV, 200 ohms, 25 µF. Use two cells for each reaction (i.e. 200 µL for each cell). After the shock, transfer cells to 25 mL SOC media and culture for phenotype expression. After phenotype expression, remove a small aliquot for titre on selective and non-selective media. Transfer cells to 500 mL 2xYT (in a 2-L baffled flask) containing appropriate

antibiotic for phagemid selection and VCS helper phage (m.o.i. = 10). Grow overnight and harvest phage in the morning.

Example 5 - Large library construction with ultrahigh DNA and *E. coli* concentrations

The fill-in protocol of Example 3 was followed with two different single stranded templates (a and b) and a mismatch oligonucleotide. Three different input template quantities were used: 1) 20 µg, 2) 30 µg, or 3) 40 µg.

After the fill-in and purification the following double stranded DNA quantities were obtained in 60 µL of water:

Reaction	OD260	[DNA], µg/mL	DNA total, µg
a1	9.04	452	27
a2	12.12	606	36
a3	13.92	696	42
b1	7.48	374	22
b2	10.4	520	31
b3	13.12	656	39

Each reaction was used to electroporate 340 µL of SS320 (3×10^{11} cfu/mL). This gives a total volume of 400 µL with 1×10^{11} cells. Each reaction was electroporated in two 200 microliter aliquots.

Reaction	[DNA], µg/mL	Time constants, ms*
a1	68	4.1, 4.1
a2	90	4.2, 4.3
a3	105	4.2, 4.3
b1	55	4.2, 4.2
b2	78	4.3, 4.3
b3	98	4.2, 4.2

*microseconds

These results indicate that the time constant for electroporations with high DNA concentrations are well above 3.0 ms and that electroporation is easily performed with high DNA concentrations.

The number of transformants and the survival of the cells during transformation was evaluated and is shown in the table below.

TITER OF LIBRARY SIZE AND SURVIVORS:

Reaction	survival	transform	S/I*	T/S**	T/I***
a1	5×10^{10}	2.3×10^{10}	0.50	0.46	0.23
a2	5×10^{10}	3.0×10^{10}	0.50	0.60	0.30
a3	4.5×10^{10}	3.0×10^{10}	0.45	0.67	0.30
b1	4×10^{10}	2.3×10^{10}	0.40	0.58	0.23
b2	4×10^{10}	2.3×10^{10}	0.40	0.58	0.23
b3	4×10^{10}	2.3×10^{10}	0.40	0.58	0.23

* survivors/input is the fraction that survive electroporation

** transformants/survivors is fraction of survivors with phagemid

***transformants/input is fraction that both survive and transform.

Example 6 – High concentration DNA electroporation

The standard fill-in protocol was followed with a single stranded template and a mismatch oligonucleotide. Seven identical reactions were purified and pooled to produce 400 µL of DNA at 750 µg/mL (OD260 = 15.0)

Various amounts of DNA were electroporated into *E. coli* SS320 in a final volume of 200 µL with a fixed *E. coli* concentration of 1.5×10^{11}cells/mL. The following conditions were used: 0.2 cm cuvettes @ 2.5 kV/cm, 200 ohms, 25 µF. After electroporation, the reaction was grown in 10 mL SOC media for 30 minutes and then titered on both LB (survival) and LB/carbenicillin (50 µg/mL) (transformation).

	[DNA] (µg/mL)	T.C. (ms)	LB (cfu)	carb (cfu)	S/I*	T/S**	T/I***
1	375	3.6	1.4×10^{10}	6.0×10^{9}	.47	.43	.20
2	281	3.8	1.4×10^{10}	7.0×10^{9}	.47	.50	.23
3	188	4.1	1.5×10^{10}	6.2×10^{9}	.50	.41	.21
4	150	4.1	1.4×10^{10}	6.0×10^{9}	.47	.43	.20
5	113	4.2	1.4×10^{10}	5.7×10^{9}	.47	.41	.19
6	75	4.3	1.5×10^{10}	5.9×10^{9}	.50	.42	.20
7	38	4.4	1.6×10^{10}	6.8×10^{9}	.53	.43	.23
8	19	4.4	1.6×10^{10}	7.0×10^{9}	.53	.44	.23
9	7.5	4.5	1.4×10^{10}	5.0×10^{9}	.47	.36	.17
10	3.8	4.5	1.5×10^{10}	2.4×10^{9}	.50	.16	.08
11	0	4.5	1.5×10^{10}	0	.50	0	0

* survivors/input is the fraction that survive electroporation

** transformants/survivors is fraction of survivors with phagemid

***transformants/input is fraction that both survive and transform.

Example 7 - Construction of an extremely large library using multiple electroporations with ultrahigh DNA and *E. coli* concentrations

The fill-in protocol of Example 3 was followed with a single stranded template and a mismatch oligonucleotide. The quantity of input template was 40 micrograms, and 35 identical reactions were performed.

After the fill-in, purification was conducted as described in Example 4 except that the extractions with phenol/$CHCl_3$ and $CHCl_3$ were omitted, and each column was eluted with 50

microL of H_2O. Two columns were used for each reaction, and thus, the final theoretical yield for each reaction was 80 micrograms of fill-in product in 100 microL of H_2O.

Each reaction was used to electroporate 700 microL of SS320 (3×10^{11} cfu/mL). This gives a total volume of 800 microL with 2×10^{11} cells. Each reaction was electroporated in two 400 microL aliquotes as described in Example 4, except that the cells were transferred to 50 mL of SOC media after the shock. After phenotype expression, the cells were titred on selective media. When combined together, the 35 independent reactions provided a library size of 1.79×10^{12} different members. The results for the 35 independent reactions are presented below:

Reaction	transformants
1	5.0×10^{10}
2	4.6×10^{10}
3	4.6×10^{10}
4	5.4×10^{10}
5	5.2×10^{10}
6	5.8×10^{10}
7	4.6×10^{10}
8	5.0×10^{10}
9	5.0×10^{10}
10	4.6×10^{10}
11	5.2×10^{10}
12	5.6×10^{10}
13	5.6×10^{10}
14	5.0×10^{10}
15	4.6×10^{10}
16	4.6×10^{10}
17	5.8×10^{10}
18	5.6×10^{10}
19	5.0×10^{10}
20	5.0×10^{10}
21	4.8×10^{10}
22	4.8×10^{10}
23	5.0×10^{10}.

24	5.2×10^{10}
25	5.6×10^{10}
26	4.6×10^{10}
27	5.4×10^{10}
28	5.4×10^{10}
29	5.8×10^{10}
30	4.6×10^{10}
31	5.0×10^{10}
32	5.4×10^{10}
33	5.8×10^{10}
34	5.0×10^{10}
35	4.8×10^{10}

Materials for further Examples

Reagents for dideoxynucleotide sequencing were from United States Biochemicals. Enzymes and plasmid pMal-p2 were from New England Biolabs. Maxisorp immunoplates were from NUNC (Roskilde, Denmark). *E. coli* XL1-Blue was from Stratagene; the construction of *E. coli* SS-320 is described above. Bovine serum albumin (BSA), Tween 20, and *o*-phenylenediamine dihydrochloride were from Sigma. HRP/anti-M13 conjugate was from Pharmacia Biotech. *Streptomyces avidinii* was from ATCC (accession no. 27419). Goat anti-streptavidin polyclonal antibody was from Zymed Laboratories (South San Francisco, USA).

Oligonucleotides for Examples 8-21

DNA degeneracies are represented in the IUB code (K = G/T, N = A/C/G/T, R = A/G, S = G/C, W = A/T, Y = C/T).

IPTG-1: AAAAGAATTCCCGACACCATCGAATGGTGC (SEQ ID NO. 9)

IPTG-2: ACCAGATGCATAAGCCGAGGCGGAAAACATCATCG (SEQ ID NO. 10)

IPTG-3: TTTTCTAGACAGGCCTCCCACCAGATGCATAAGCCGAGGCGGAAAA CATCATCGTC (SEQ ID NO. 11)

SAV-1: GGCTATCGGAATGCATCGGGCATCACCGGCACCTG (SEQ ID NO. 12)

SAV-2: GAGTCATAGTCGTCAGGCGCCTCCTCCGGATCCTCCACCCACCTT GGTGAAGGTGTCGTGG (SEQ ID NO. 13)

hGH-1: GGGTATCTAGAGGTTGAG (SEQ ID NO. 14)

hGH-2: TGGAGCTCCCGGATCCTCCACCGCTCTGGAAGCCACAGCTGCCCTC (SEQ ID NO. 15)

g8stop1: GGATCCGGGAGCTCCAGCTGATGAGGTGACGATCCCGCAAAA (SEQ ID NO. 16)

g8stop2: GATCCCGCAAAAGCGGCCTGATGATCCCTGCAAGCCTCAGCG (SEQ ID NO. 17)

g8stop3: CAAGCCTCAGCGACCGAATGATGAGGTTATGCGTGGGCGATG (SEQ ID NO. 18)

g8stop4: GCGTGGGCGATGGTTGTTTGATGAGTCGGCGCAACTATCGGT (SEQ ID NO. 19)

g8stop5: GCAACTATCGGTATCAAGTGATGAAAGAAATTCACCTCGAAA (SEQ ID NO. 20)

g8V1: GGATCCGGGAGCTCCAGCRNTNASRNTNASNASNYCRNTRNARNT RNTTTTAACTCCCTGCAAGCC (SEQ ID NO. 21)

g8V2: GATCCCGCAAAAGCGGCCNWTNASRNTNYTNASRNTRNTRNTRNTN ASTATATCGGTTATGCGTGG (SEQ ID NO. 22)

g8V3: CAAGCCTCAGCGACCGAANWCNWCNKTNWCNYYTNKGNYTNKGN WTNWTGTCATTGTCGGCGCAACTATC (SEQ ID NO. 23)

g8V4: GCGTGGGCGATGGTTGTTNWTNWCNWTNKTNYTNYTNNTNNTNNT AAGCTGTTTAAGAAATTCACC (SEQ ID NO. 24)

g8V5: GCAACTATCGGTATCAAGNNGNNSAAGAAANNSNNGNNGAAANN GNNGTGATAAACCGATACAATTAAAGGC (SEQ ID NO. 25)

g8(1a): GATCCCGCAAAAGCGGCCTATGAGGCTCTTGAGGATATTGCTAC TAACTATATCGGTTATGCGTGG (SEQ ID NO. 26)

R64A: CCGACACCCTCCAATGCTGAGGAAACACAACAGAAA (SEQ ID NO. 27)

D171A: TTCAGGAAGGACATGGCTAAGGTCGAGACATTCCTG (SEQ ID NO. 28)

Y164A/R178A: AACTACGGGCTGCTCGCTTGCTTCAGGAAGGACATGGACAAGG TCGAGACATTCCTGGCTATCGTGCAGTGCCGC (SEQ ID NO. 29)

K172A/R178A: TTCAGGAAGGACATGGACGCTGTCGAGACATTCCTG GCTATCGTCCAGTGCCGCTCT (SEQ ID NO. 30)

Lstop: GGTGGAGGATCCGGGAGCTGATGAGCCGAGGGTGACGATCCC (SEQ ID NO. 31)

Lstop2: CACCAAGGTGGTCTAGAGCTAATAATAAGCCGAGGGTGACGA TCCC (SEQ ID NO. 32)

LV: GAGGGCAGCTGTGGCTTCGGTGGCGGTVVCVVCVVCVVCVVCV (SEQ ID NO. 33) VCVVCVVCVVCVVCVVCVVCVVCVVCGGCGGTGCCGAGGGTGACGATCCC (SEQ ID NO. 34)

LV5: CACCAAGGTGGTCTAGAGCVVCVVCVVCVVCVVCGCCGAGGGTGA CGATCCC (SEQ ID NO. 35)

LV10: CACCAAGGTGGTCTAGAGCVVCVVCVVCVVCVVCVVCVVCVV CVVCVVCGCCGAGGGTGACGATCCC (SEQ ID NO. 36)

LV15: CAC CAAGGTGGTCTAGAGCVVCVVCVVCVVCVVCVVCVVCVV
CVVCVVCVVCVVCVVCVVCVVCGCCGAGGGTGACGATCCC (SEQ ID NO. 37)
LV20: CACCAAGGTGGTCTAGAGCVVCVVCVVCVVCVVCVVCVVCVV
CVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCGCCGAGGGTGA CGATCCC
(SEQ ID NO. 38)
LV25: CACCAAGGTGGTCTAGAGCVVCVVCVVCVVCVVCVVCVVCVV
CVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVVCVV
CVVCVVCGCCGAGGGTGACGATCCC (SEQ ID NO. 39)
g8V2c: AAGTTCGCTAGAGATGCTTATGAGGCTCTTGAGGATATTGCTAC
TAACTATATCGGTTATGCGTGG (SEQ ID NO. 40)
g8V3c: GAGGATATTGCTACTAACCTTTTCTTTCTCCTTGGGACTGTGC
ATCTTGTCATTGTCGGCGCAACT (SEQ ID NO. 41)
g8V2-E12N: GCAAAAGCGGCCTATAACGCTCTTGAGGATATT (SEQ ID NO. 42)
g8V2-D16A: TATGAGGCTCTTGAGGCCATTGCTACTAACTAT (SEQ ID NO. 43)
g8V2-I17S: GAGGCTCTTGAGGATTCAGCTACTAACTATATC (SEQ ID NO. 44)
g8V2b: GATCCCGCAAAAGCGGCCTATGAGGCTCTTGAGGATATTGCTA
CTAACTATATCGGTTATGCGTGG (SEQ ID NO. 45)
L-wt: GAGGGCAGCTGTGGCTTCCAGAGCGGTGGAGGATCCGGGAG
CTCCAGCGCCGAGGGTGACGATCCC (SEQ ID NO. 46)
S13A/S17I: CCCGCAAAAGCGGCCTTTAACGCTCTGCAAGCCATTGCGACC
GAATATATCGGTTATGCG (SEQ ID NO. 47)
g8V3b: CAAGCCTCAGCGACCGAACTTTTCTTTCTCCTTGGGACTGTGCATC
TTGTCATTGTCGGCGCAACT (SEQ ID NO. 48)
D1A: TCCGGGAGCTCCAGCGCCAAGAGTGAGAAGTTC (SEQ ID NO. 49)
K2E: GGGAGCTCCAGCGATGAGAGTGAGAAGTTCGCT (SEQ ID NO. 50)
S3G: AGCTCCAGCGATAAGGGTGAGAAGTTCGCTAGA (SEQ ID NO. 51)
E4D: TCCAGCGATAAGAGTGACAAGTTCGCTAGAGAT (SEQ ID NO. 52)
K5D: AGCGATAAGAGTGAGGATTTCGCTAGAGATGCT (SEQ ID NO. 53)
F6P: GATAAGAGTGAGAAGCCCGCTAGAGATGCTTTT (SEQ ID NO. 54)
R8K: AGTGAGAAGTTCGCTAAAGATGCTTTTAACTCC (SEQ ID NO. 55)
D9A: GAGAAGTTCGCTAGAGCGGCTTTTAACTCCCTG (SEQ ID NO. 56)
Y11F: CCCGCAAAAGCGGCCTTTGAGGCTCTTGAGGAT (SEQ ID NO. 57)
E12N: GCAAAAGCGGCCTATAAACGCTCTTGAGGATATT (SEQ ID NO. 58)
A13S: AAAGCGGCCTATGAGTCCCTTGAGGATATTGCT (SEQ ID NO. 59)
E15Q: GCCTATGAGGCTCTTCAAGATATTGCTACTAAC (SEQ ID NO. 60)
D16A: TATGAGGCTCTTGAGGCCATTGCTACTAACTAT (SEQ ID NO. 61)
I17S: GAGGCTCTTGAGGATTCAGCTACTAACTATATC (SEQ ID NO. 62)
N20E: GAGGATATTGCTACTGAATATATCGGTTATGCG (SEQ ID NO. 63)

L21Y: GCCTCAGCGACCGAATATTTCTTTCTCCTTGGG (SEQ ID NO. 64)
F22I: TCAGCGACCGAACTTATCTTTCTCCTTGGGACT (SEQ ID NO. 65)
F23G: GCGACCGAACTTTTCGGTCTCCTTGGGACTGTG (SEQ ID NO. 66)
L24Y: ACCGAACTTTTCTTTTATCTTGGGACTGTGCAT (SEQ ID NO. 67)
L25A: GAACTTTTCTTTCTCGCGGGGACTGTGCATCTT (SEQ ID NO. 68)
G26W: CTTTTCTTTCTCCTTTGGACTGTGCATCTTGTC (SEQ ID NO. 69)
T27A: TTCTTTCTCCTTGGGGCGGTGCATCTTGTCATT (SEQ ID NO. 70)
V28M: TTTCTCCTTGGGACTATGCATCTTGTCATTGTC (SEQ ID NO. 71)
H29V: CTCCTTGGGACTGTGGTTCTTGTCATTGTCGGC (SEQ ID NO. 72)
L30V: CTTGGGACTGTGCATGTTGTCATTGTCGGCGCA (SEQ ID NO. 73)
E12N/A13S: GCAAAAGCGGCCTATAACTCCCTTGAGGATATTGCT (SEQ ID NO. 74)
E12N/I17S: GCAAAAGCGGCCTATAACGCTCTTGAGGATTCAGCTACT
AACTATATC (SEQ ID NO. 75)
A13S/I17S: CCCGCAAAAGCGGCCTATGAGTCCCTTGAGGATTCAGCTACT
AACTATATCGGTTATGCG (SEQ ID NO. 76)
E12N/A13S/I17S: GCAAAAGCGGCCTATAACTCCCTTGAGGATTCAGCT
ACTAACTATATC (SEQ ID NO. 77)

Example 8 - pS349: A phagemid for phage display of hGH on Protein VIII.

A DNA fragment containing the gene for hGH was amplified using the PCR (with a derivative of plasmid pB0475 (Cunningham, B. C., Jhurani, P., Ng, P., and Wells, J. A. (1989) *Science* 243:1330-1336) as template and oligonucleotides *hGH-1* and *hGH-2* as primers). The DNA fragment was digested with *Nsi*I and cloned into a protein VIII display phagemid (Lowman, H. B., Chen Y. M., Skelton, N. J., Mortensen, D. L., Tomlinson, E. E., Sadick, M. D., Robinson, I. C. A. F., and Clark, R. G. (1998) *Biochemistry* 37:8870-8878) which had been first digested with *Kas*I and treated with T4 DNA polymerase to produce blunt ends and subsequently digested with *Nsi*I. The resulting phagemid was designated pS135a.

The PCR was used with primers *IPTG-1* and *IPTG-2* to amplify a 1.6 kbp fragment of pMal-p2 (New England Biolabs Product Catalog (1996-97) p212) containing the *lacI*^{q} gene and a gene fragment encoding the signal peptide from maltose binding protein under the control of the P_{tac} promoter. The DNA fragment was digested with *Eco*RI and *Nsi*I and ligated with the large fragment resulting from a similar digestion of pS135a. The resulting phagemid (designated pS349) contains a gene encoding a fusion product (the maltose binding protein signal peptide, followed by hGH, followed by a Gly/Ser-rich linker peptide (QSGGGSGSSS) (SEQ ID NO. 78), and protein VIII of *E. coli* bacteriophage M13 under the control of the IPTG-inducible P_{tac} promoter (New England Biolabs). In addition, pS349 also contains the *lacI*^{q} gene for effective transcription repression in the absence of IPTG.

Example 9 - pW277e: A phagemid for phage display of SAV on protein VIII.

A derivative of pS349 was constructed and designated pS657a. pS657a differs from pS349 in two respects. Firstly, the gene encoding hGH has been replaced by sequence encoding a pentapeptide (GGRPV) (SEQ ID NO. 79). Secondly, the introduction of an *Xba*I site in the linker preceding protein VIII has changed the codon encoding glutamine to an amber (TAG) stop codon. Digestion with *Nsi*I and *Xba*I excises the pentapeptide-encoding sequence and allows for the directional cloning of appropriately digested DNA fragments into a position analogous to that of the hGH gene in pS349.

A PCR was performed with *Streptomyces avidinii* genomic DNA as template and oligonucleotides *SAV-1* and *SAV-2* as primers. The amplified DNA fragment contained codons 16 to 133 of the streptavidin (SAV) gene open reading frame flanked by an *Nsi*I site at the 5' end and an *Xba*I site at the 3' end. The fragment was digested with *Nsi*I and *Xba*I and cloned into similarly digested phagemid pS657a. The resulting phagemid (pW277e) encodes a fusion product similar to that encoded by pS349 except that hGH has been replaced by SAV. Also, an amber codon has been positioned between the segments encoding SAV and protein VIII.

Example 10 - Construction of mutant protein VIII libraries.

For library construction, protein VIII was divided into five zones encompassing approximately 10 contiguous residues each (zone 1, residues 1 to 10; zone 2, residues 11 to 20; zone 3, residues 21 to 30; zone 4, residues 31 to 39; zone 5, residues 40 to 50). Libraries were constructed using a modified version (SS320 described above) of a previously described method (Lowman, H. B. (1998) Phage Display of Peptide Libraries on Protein Scaffolds. From: *Methods in Molecular Biology, vol. 87: Combinatorial Peptide Library Protocols*. Edited by: S. Cabilly. Publisher: Humana Press Inc., Totowa, NJ). Briefly for each zone, an oligonucleotide (*g8stopn*, where "*n*" is the zone number) was used with either pS349 (for hGH display) or pW277e (for SAV display) as template to introduce two consecutive TGA stop codons within the zone, using the method of Kunkel (Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* 82:488-492). The resulting phagemid was used as template in a second round of the Kunkel method with a degenerate oligonucleotide (*g8Vn*, where "*n*" is the zone number) designed to introduce mutations at the desired sites.

Libraries were constructed for each zone of the protein VIII moiety within the hGH-protein VIII fusion product of pS349. The diversities of these libraries were as follows: zone-1, 2.5×10^{10}; zone-2, 2.5×10^{10}; zone-3, 2.5×10^{10}; zone-4, 1.3×10^{10}; and zone-5, 5.0×10^{9}. Libraries were constructed for zone-1, zone-2, and zone-3 of the protein VIII moiety within the SAV- protein VIII fusion product of pW277e. The diversities of these libraries were as follows: zone-1, 3.0×10^{9}; zone-2, 6.8×10^{9}; and zone-3, 8.6×10^{9}.

Libraries were also constructed to vary the linker between the fused protein and protein VIII. For hGH display, the oligonucleotide *Lstop* was used to introduce two consecutive TGA stop

codons within the linker between hGH and protein VIII. The resulting phagemid was used as template in a second round of the Kunkel method with a degenerate oligonucleotide (*LV*) designed to introduce a linker of the form $(Gly)_3(Xaa)_{14}(Gly)_2$ (where Xaa is a variable position) in place of the Gly/Ser-rich linker encoded by pS349. For SAV display, the linker was varied between SAV and variant protein VIII (1e) (see Figure 2, for the sequence of protein VIII (1e)). The oligonucleotide *Lstop2* was used to introduce three consecutive TAA stop codons within the linker between SAV and protein VIII (1e). The resulting phagemid was used as template for the production of libraries with linkers of variable length and sequence. The oligonucleotides *LV5, LV10, LV15, LV20,* and *LV25* were used to construct libraries with linkers containing 5, 10, 15, 20, or 25 variable residues, respectively.

The diversities of the linker libraries were as follows: hGH-LV- protein VIII, 1.8×10^{10}; SAV-LV5- protein VIII, 1.4×10^{10}; SAV-LV10- protein VIII, 9.8×10^{9}; SAV-LV15- protein VIII, 1.2×10^{10}; SAV-LV20- protein VIII, 1.1×10^{10}; SAV-LV25- protein VIII, 6.0×10^{9}.

Example 11 - Selection of protein VIII variants which increase fusion protein display.

Phage from the hGH- protein VIII libraries described above were cycled through rounds of binding selection with hGHbp (Fuh, G. *et al.* (1990) *J. Biol. Chem.* 265:3111; Cunningham, B. C., Ultsch, M., De Vos, A. M., Mulkerrin, M. G., Clauser, K. R., and Wells, J. A. (1991) *Science* 254:821-825) coated on 96-well Maxisorp immunoplates as a target. All libraries were sorted separately. Phage were propagated in *E. coli* SS320 cells with M13-VCS helper phage (Stratagene). After five rounds of binding selection, individual phage were isolated and analyzed for hGH display using a phage ELISA with hGHbp as target (see below). Phage exhibiting strong signals in the phage ELISA were sequenced (Sanger, F. *et al.* (1979) *Proc. Natl. Acad.Sci. USA,* 74:5463-5467).

The SAV- protein VIII libraries were pooled and binding selection was performed as described above for the hGH- protein VIII libraries, except that the binding target was an anti-SAV polyclonal antibody. Phage were propagated in the *SupE E. coli* strain XL1-Blue in which the amber stop codon is suppressed as glutamine (Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) *Biotechniques* 5:376-379).

Example 12 - Site-directed mutagenesis.

Mutageneses were performed using the method of Kunkel (Kunkel, *et al.* (1987) Meth. Enzymol. 154:367-382). Template DNA was prepared by growing an appropriate plasmid (e.g., containing the hGH gene fused to the carboxy-terminal half of M13 gene III) in host cells with M13-K07 phage added as helper phage. Single-stranded, uracil-containing DNA was prepared for mutagenesis to introduce the desired mutation into the hGH-protein VIII gene fusion. Oligonucleotide-directed mutagenesis was carried out using T7 DNA polymerase and the

appropriate oligodeoxy-nucleotides. Clones from the mutagenesis were confirmed by dideoxy DNA sequencing.

The mutagenic oligonucleotide *g8(1a)* was used to introduce the protein VIII mutations of selectant hGH- protein VIII (1a) into phagemid pW277e. Mutant hGH genes were constructed using the appropriately named oligonucleotide (*e.g.*, oligonucleotide *R64A* encodes the mutation of Arg64 to Ala).

The mutagenic oligonucleotide *g8V2c* was used to introduce the mutations of protein VIII(2a) into the gene encoding protein VIII(1a). The mutagenic oligonucleotide *g8V3c* was used to introduce the mutations of protein VIII(3a) into the gene encoding protein VIII(2a). The mutation E12N, D16A, or I17S was introduced into the gene encoding protein VIII(2a) using the mutagenic oligonucleotide *g8V2-E12N*, *g8V2-D16A*, or *g8V2-I17S*.

Example 13 - Phage ELISAs for determining relative levels of fusion protein display on protein VIII and protein VIII variants thereof.

Cultures of *E. coli* XL1-Blue (Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) *Biotechniques* 5:376-379) harboring phagemids were grown for 8 hours at 37°C in 1 mL of 2YT, 50 μg/mL carbenicillin, 10 μg/mL tetracycline. The cultures were transferred to 30 mL of the same media (supplemented with M13-VCS helper phage (10^{10} phage/mL) and IPTG at the appropriate concentration) for overnight growth at 37°C. Phage were harvested from the culture supernatant by precipitation twice with PEG/NaCl (Lowman, H. B., (1998) Phage Display of Peptide Libraries on Protein Scaffolds. From: *Methods in Molecular Biology, vol. 87: Combinatorial Peptide Library Protocols*. Edited by: S. Cabilly. Publisher: Humana Press Inc., Totowa, NJ) and resuspended in 1 mL of PBS, 0.2% BSA, 0.1% Tween (BSA blocking buffer). Phage concentrations were determined spectrophotometrically ($\epsilon 268 = 1.2 \times 10^8\ M^{-1}cm^{-1}$).

Maxisorp immunoplates (96-well) were coated with target protein for 2 hours at room temperature (100 μL at 5 μg/mL in 50 mM carbonate buffer, pH 9.6). The plates were then blocked for 1 h with 0.2% BSA in phosphate-buffered saline (PBS) and washed (8X) with PBS, 0.05% Tween 20. Phage particles were diluted serially using PBS, 0.2% BSA, 0.1% Tween (BSA blocking buffer) and then transferred (100 μL) to coated wells. After 1 h, plates were washed (8X) with PBS, 0.05% Tween 20, incubated with 100 μL of 1:3000 HRP/anti-M13 conjugate in BSA blocking buffer for 30 min, and then washed with PBS, 0.05% Tween 20 (8X) and PBS (2X). Plates were developed using an *o*-phenylenediamine dihydrochloride/H_2O_2 solution (100 μL), stopped with 2.5 M H_2SO_4 (50 μL), and read spectrophotometrically at 492 nm. Li, B. et al., (1995) *Science* 270:165-1660.

Example 14 - display of SAV on a protein VIII variant selected for increased display of hGH.

The method of Kunkel (Example 12) was used with the mutagenic oligonucleotide *g8V2b* to introduce the protein VIII mutations of selectant hGH-protein VIII(2a) into phagemid pW277e. The resulting phagemid encoded a fusion protein identical to that encoded by pW277e except that the protein VIII within the fusion moiety contained the mutations of variant protein VIII(2a). SAV display was measured by phage ELISAs (Example 13) with either anti-SAV polyclonal antibody (Fig. 4a) or biotinylated BSA (Fig. 4b) as target.

Example 15 - display and detection of hGH variants with attenuated binding affinities.

hGH mutants with reduced site 1 binding affinity for hGHbp (Pierce et al., above) were displayed as fusions to either wild type protein VIII or variant protein VIII(1a). Mutant hGH genes were constructed using the method of Kunkel (Example 12) with the following mutagenic oligonucleotides: hGH(R64A), oligonucleotide *R64A*; hGH(D171A), oligonucleotide *D171A*; hGH(Y164A/R178A), oligonucleotide *Y164A/R178A*; hGH(K172A/R178A), oligonucleotide *K172A/R178A*. For display of hGH fused with wild type protein VIII, the mutagenesis template was pS349. For display of hGH fused with variant protein VIII(1a), the template was a derivative of pS349 encoding a fusion protein consisting of hGH fused to protein VIII(1a). hGH display was measured by phage ELISA (Example 13) with the hGHbp as target (Fig. 3).

Example 16 - display of hGH with protein VIII variants combining mutations in different zones.

The method of Kunkel (Example 12) was used to combine mutations from protein VIII variants independently selected for increased hGH display. The oligonucleotide *g8V2c* was used to introduce the mutations of protein VIII(2a) into the gene encoding protein VIII(1a). The oligonucleotide *g8V3c* was used to introduce the mutations of protein VIII(3a) into the gene encoding protein VIII(2a). hGH display was measured by phage ELISA (Example 13) with an anti-hGH monoclonal antibody as target (Fig. 5).

Example 17 - display of hGH with protein VIII variants derived from protein VIII(2a) through back mutations to the wild type sequence.

The method of Kunkel (Example 12) was used to introduce the mutation E12N, D16A, or I17S into the gene encoding protein VIII(2a) using the mutagenic oligonucleotide *g8V2-E12N, g8V2-D16A,* or *g8V2-I17S*, respectively. hGH display was measured by phage ELISA (Example 13) with an anti-hGH monoclonal antibody as target (Fig. 6).

Example 18 - display of hGH with phagemid pS1607.

Further sequence analysis revealed that the fusion-protein VIII gene contained within the pS349 clone used in Figures 2, 3, 5, and 8 had a deletion of five base pairs consisting of the final base pair encoding the linker between hGH and protein VIII and the first four base pairs encoding

protein VIII. This deletion introduced a frameshift which reduced hGH display. The method of Kunkel (Example 12) was used with the mutagenic oligonucleotide *L-wt* to correct the frameshift in pS349. The resulting phagemid was designated pS1607. Phagemid pS1607 differs from pS349 only in the addition of five base pairs to correct this frameshift. In pS349, the sequence following the hGH gene is as follows: CAGAGCGGTGGAGGATCCGGGAGCTCCAGAGGGT (the underlined bases are part of the beginning of the protein VIII gene) (SEQ ID NO. 80). In pS1607, the corresponding sequence is as follows: CAGAGCGGTGGAGGATCCGGGAGCTCCAgcgccGAGGGT (lower case indicates the bases inserted through mutagenisis with oligonucleotide L-wt) (SEQ ID NO. 81). hGH display was measured by phage ELISA (Example 13) with the hGHbp as target (Fig. 9).

Example 19 - Fab display with a protein VIII variant.

Phagemid pS1705a directs secretion of a free Fab light-chain and a Fab heavy-chain fused to protein VIII. The Fab heavy-chain also contains a peptide flag (MADPNRFRGKDL) (SEQ ID NO. 82) fused to its N-terminus which can be detected with a specific monoclonal antibody. The method of Kunkel (Example 12) was used with the mutagenic oligonucleotide *S13A/S17I* and the template pS1705a, the resulting phagemid was designated pS1709b. Phagemid pS1709b is identical to pS1705a except that the protein VIII gene linked to the Fab heavy-chain contains the mutations S13A/S17I which were observed in protein VIII variants selected for increased display of hGH (Fig. 1B). Fab display was measured by phage ELISA (Example 13) with a peptide flag-specific monoclonal antibody as target. Fab display with pS1709b was greater than Fab display with pS1705a (Fig. 10). Thus protein VIII mutations which were selected for increased display of hGH also increased Fab display.

Example 20 - display of hGH with protein VIII variants combining mutations in different zones.

The method of Kunkel (Example 12) was used to combine mutations from protein VIII variants independently selected for increased hGH display. The oligonucleotide *g8V2c* was used to introduce the mutations of protein VIII(2a) into the gene encoding protein VIII(1a). The oligonucleotide *g8V3c* was used to introduce the mutations of protein VIII(3a) into the gene encoding protein VIII(2a). The oligonucleotide *g8V3b* was used to introduce the mutations of protein VIII(3a) into the gene encoding protein VIII(1a). The oligonucleotide *g8V3c* was used to introduce the mutations of protein VIII(3a) into the gene encoding protein VIII containing the mutations of protein VIII(1a) and protein VIII(2a). hGH display was measured by phage ELISA (Example 13) with an anti-hGH monoclonal antibody as target. All protein VIII variants increased hGH display in comparison with wild-type protein VIII (Fig. 11).

Example 21 - modulated display of hGH with protein VIII variants derived from protein VIII selectants through back mutations to the wild-type sequence.

Back mutation scanning is the independent conversion of each mutation within a coat protein, such as protein VIII, variant back to the wild-type sequence. A protein VIII selectant with mutations in either zone 1, 2, or 3 was subjected to back mutation scanning analysis. The following selectants were analyzed: protein VIII(1a), protein VIII(2a), and protein VIII(3a) (Fig. 1). The method of Kunkel (Example 12) was used to mutate each mutation in a given variant back to the wild-type sequence. Appropriately designed and named oligonucleotides were used (*e.g.* the oligonucleotide *D1A* mutates Asp1 in protein VIII(1a) to Ala). In addition, double and triple back mutations were introduced into protein VIII(2a), again using appropriately designed and named oligonucleotides (*e.g.*, the oligonucleotide *A13S/I17S* simultaneously mutates A13 and I17 to Ser).

hGH display was measured by phage ELISA (Example 13) with an anti-hGH monoclonal antibody as target. Some of the back mutations reduced hGH display, allowing for the modulation of hGH display (Fig. 12).

Oligonucleotides for Examples 22-25

Add-NX: GATGGTGAAGCTGCGGCTGATGCATCTGGTAGCGTCTAGAGC CACCATCACCATCACCAT (SEQ ID NO: 83)

add-P12-7: GCTGTCGGTATTATTTACATGCTCCTCGTGGAGGCGTCGCCC TGGGCTGCTAAGGCGCCA (SEQ ID NO: 84)

G-0: ACCTCGAAAGCAAGCCATCACCATCACCATGCG (SEQ ID NO: 85)

G-1: ACCTCGAAAGCAAGCGGCCATCACCATCACCATGCG (SEQ ID NO: 86)

G-2: ACCTCGAAAGCAAGCGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 87)

G-3: ACCTCGAAAGCAAGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 88)

G-4: ACCTCGAAAGCAAGCGGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 89)

G-6: ACCTCGAAAGCAAGCGGTGGTGGCGGTGGTGGCCATCAC CATCACCATGCG (SEQ ID NO: 90)

G-7: ACCTCGAAAGCAAGCGGCGGTGGTGGCGGTGGTGGCCATCAC CATCACCATGCG (SEQ ID NO: 91)

G8: ACCTCGAAAGCAAGCGGTGGCGGTGGTGGCGGTGGTGGCCATCACCATCACCAT GCG (SEQ ID NO: 92)

G-9: ACCTCGAAAGCAAGCGGCGGTGGCGGTGGTGGCGGTGGTGGCCATCAC CATCACCATGCG (SEQ ID NO: 93)

G-10: ACCTCGAAAGCAAGCGGTGGCGGTGGCGGTGGTGGCGGTGGTGGC CATCACCATCACCATGCG (SEQ ID NO: 94)

G-12: ACCTCGAAAGCAAGCGGTGGCGGTGGCGGTGGCGGTGGTGGCGGT GGTGGCCATCACCATCACCATGCG (SEQ ID NO: 95)

G-14: ACCTCGAAAGCAAGCGGTGGTGGTGGCGGTGGCGGTGGCGGT
GGTGGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 96)
G-16: ACCTCGAAAGCAAGCGGCGGCGGTGGTGGTGGCGGTGGCGGTGGC
GGTGGTGGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 97)
G-18: ACCTCGAAAGCAAGCGGCGGTGGCGGCGGTGGTGGTGGCGGTGGC
GGTGGCGGTGGTGGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 98)
G-20: ACCTCGAAAGCAAGCGGTGGTGGCGGTGGCGGCGGTGGTGGTGGC
GGTGGCGGTGGCGGTGGTGGCGGTGGTGGCCATCACCATCACCATGCG (SEQ ID NO: 99)
Lib-zone1: CAAGGACCATAGATTATGNNSNNSNNSNNSNNSNNSAAGTT
TCTGAAAGTTTTTGTTTTT (SEQ ID NO: 100)
Lib-zone2: ATTATGAGCAAGAGCACTNNSNNSNNSNNSNNSNNSGTT
TTTGTTTTTTCTGTTGAT (SEQ ID NO: 101)
Lib-zone3: TTCAAAAAGTTTCTGAAANNSNNSNNSNNSNNSNNSNNS
NNSNNSAATTGGATTTGGGCTGTCGGT (SEQ ID NO: 102)
Lib-zone4: GTTTTTTCTGTTGATGTTGATNNSNNSNNSNNSNNSNNSNNS
NNSNNSNNSGCGGCTGATGCATTCCCA (SEQ ID NO: 103)
Lib-zone5: TGGGCTGTCGGTATTATTNNSNNSNNSNNSNNSNNSNNS
NNSNNSGCTGCTAAGGCGCCAGACGATGGT (SEQ ID NO: 104)
Lib-zone6: AGCGCTCAGCTGAGCAACTTCNNSNNSNNSNNSNNSNNS
NNSNNSNNSGCGGCTGATGCATTCCCA (SEQ ID NO: 105)
Lib-linker: GATGGTGAAGCTGCGGCTVVCVVCVVCVVCVVCVVC
VVCVVCVVCVVCVVCVVCVVCFATGCATTCCCAACTATACCA (SEQ ID NO: 106)
Pep-ins: ACTTTCAAAAAGTTTCTGAAANWTNKTNWTNYTNYTNKT
NWTNWTNWTNWTNWTNKGNYTNKGNYTNWCNKTNWTNWTGAGACTGCTAGCGCTC
AG (SEQ ID NO: 107)

Example 22 - Determination of optimal linker length for the display of a peptide fused to the C-terminus of protein VIII.

Standard molecular biology techniques were used to construct a phagemid designated pS1290a. pS1290a is identical to phagemid pS349 (see Example 8) except that the open reading frame (ORF) under the control of the IPTG-inducible Ptac promoter (New England Biolabs) has been deleted and replaced by a new ORF. The new ORF encodes a fusion product consisting of the maltose binding protein signal peptide, followed by a Ser residue, followed by residues 2-50 of mature protein VIII of *E. coli* bacteriophage M13. The ORF is followed by two TAA stop codons, followed by sequence (CACCATCACCATCACCATGCG) (SEQ ID NO: 108) encoding a heptapeptide (HHHHHHA, hexaHis) flag or epitope tag, followed by two stop codons (TGATAA).

pS1290a was mutated using the method of Kunkel (Example 12). The two TAA stop codons and the first His codon following the protein VIII C-terminus were replaced by various numbers of Gly codons. Appropriately designed and named mutagenic oligonucleotides were used (*e.g.*, oligonucleotide *G-6* inserts six Gly codons). This resulted in the construction of a series of phagemids encoding ORFs designed to secrete protein VIII molecules with C-terminal fusions consisting of linkers containing varying numbers of Gly residues followed by a pentaHis flag (HHHHHA). The number of Gly residues was varied from zero (*i.e.*, the polyHis flag was fused directly to the protein VIII C-terminus) to 20. PentaHis flag display was measured by phage ELISA (Example 13) with an anti-(His)5 antibody (Qiagen) as the capture target (Fig. 13).

Example 23 - Optimization of the linker sequence for display of a peptide fused to the C-terminus of protein VIII.

Libraries were constructed to vary the linker between protein VIII and the hexaHis flag encoded by pS1290a. Libraries were constructed using a modified version of a previously described method (see Example 10). Mutagenic oligonucleotides were used to replace the two TAA stop codons between protein VIII and the hexaHis flag with libraries of linkers. The lengths of the linkers were varied and depended on the mutagenic oligonucleotides used: oligonucleotides *UH-L4*, *UH-L5*, *UH-L6*, *UH-L8*, or *UH-L10* introduced linkers containing 4, 5, 6, 8, or 10 residues respectively. The total diversity of the linker libraries was 5.7×10^{10}.

Phage from the linker libraries described above were pooled together and cycled through rounds of binding selection with an anti-(His)4 antibody (Qiagen) as the capture target. After two rounds of selection, individual clones were assayed for hexaHis flag display using a phage ELISA with the anti-(His)4 antibody as target. Clones exhibiting the strongest signals were subjected to DNA sequence analysis and the linker sequences were deduced from the DNA sequence and are shown below.

GCC	TGG	GAG	GAG	AAC	ATC	GAC	AGC	GCC	CCC	(SEQ ID NO. 109)
A	W	E	E	N	I	D	S	A	P	(SEQ ID NO. 110)
CAG	TAC	GGG	ACG	CCG	GAC	ACC	GAC	ACC	GAC	(SEQ ID NO. 111)
Q	Y	G	T	P	D	T	D	T	D	(SEQ ID NO. 112)
ACG	GGG	TGG	TTG	GAG	GGG	CCC	GAC	ACC	CCC	(SEQ ID NO. 113)
T	G	W	L	E	G	P	D	T	P	(SEQ ID NO. 114)
CTC	ATG	GGC	CCC	GGC	GCG	GAC	GGC			(SEQ ID NO. 115)
L	M	G	P	G	A	D	G			(SEQ ID NO. 116)
CAC	GAC	TCG	GTC	CCG	AGC	AAC	GGC			(SEQ ID NO. 117)
H	D	S	V	P	S	N	G			(SEQ ID NO. 118)

Linkers selected for display of a peptide fused to the C-terminus of protein VIII. The sequences shown were inserted between the final residue of protein VIII and a heptapeptide (HHHHHHA, referred to as a hexaHis flag). For each selectant, the DNA sequence is shown with

the deduced amino acid sequence below. The numerical designation for each sequence is shown to the left.

The level of polyHis flag display achieved with an optimized linker was compared with levels of display achieved with poly-glycine linkers of various length (Example 22) using a phage ELISA with an anti-(His)5 antibody (Qiagen) as the capture target (Fig. 14).

Example 24 – Design and selection of a new phage coat protein (Protein-12, P12) for the display of polypeptides as C-terminal fusions.

This example demonstrates the de novo design of a phage coat protein and the display of a protein of interest on the surface of phage particles containing the fusion protein, illustrating the broad scope of the method of the invention to prepare any variant phage coat protein. The retrotranslation of a peptide is the backward reading of the primary sequence and the resulting peptide is the retro-peptide of the original peptide. For example, the retrotranslation of the peptide Gly-Ala-Leu is the retro-peptide Leu-Ala-Gly.

Standard molecular biology techniques were used to construct a phagemid designated pS1207a. pS1207a is identical to phagemid pS349 (Example 8) except that the ORF under the control of the IPTG-inducible Ptac promoter (New England Biolabs) has been deleted and replaced by a new ORF. The DNA sequence of the new ORF is as follows:

ATGAGCAAGAGCACTTTCAAAAAGTTTCTGAAAGAGACTGCTAGCGCTCAGCTGAGC
AACTTCGCTGCTAAGGCGCCAGACGATGGTGAAGCTGCGGCTCACCATCACCATCACC
ATGCG (SEQ ID NO: 119)

The new ORF encodes the following polypeptide:

MSKSTFKKFLKETASAQLSNFAAKAPDDGEAAAHHHHHHA. (SEQ ID NO: 120).

This ORF was designed as follows. The first two residues were (Met-Ser) chosen to allow good translation initiation. This dipeptide was followed by a retrotranslation of residues 40-48 of mature protein VIII from M13 bacteriophage (KLFKKFTSK retrotranslated to KSTFKKFLK) which was in turn followed by a retrotranslation of protein VIII residues 1-20 (AEGDDPAKAAFNSLQASATE retrotranslated to ETASAQLSNSAAKAPDDGEA). To the C-terminus of this polypeptide was fused a nonapeptide (AAHHHHHHA) hexaHis flag. Thus, this ORF consists of the dipeptide Met-Ser, followed by a retrotranslation of residues 1-48 of mature protein VIII with the central hydrophobic section (residues 21-30) deleted, followed by a hexaHis flag.

A library of 19-mer peptides was inserted between residues 11 and 12 of the above described ORF, using a modified version of a previously described method (see Example 10) with pS1207a as the template and *Pep-ins* as the mutagenic oligonucleotide. The resulting library encoded ORFs with the following sequences:

MSKSTFKKFLK-(x)19-ETASAQLSNFAAKAPDDGEAAAHHHHHHA (SEQ ID NO: 121)

where "(x)19" indicates a random 19-mer peptide library. The degenerate codons used at each position within the library are shown in Figure 15. The library diversity was 8.3×10^{10}.

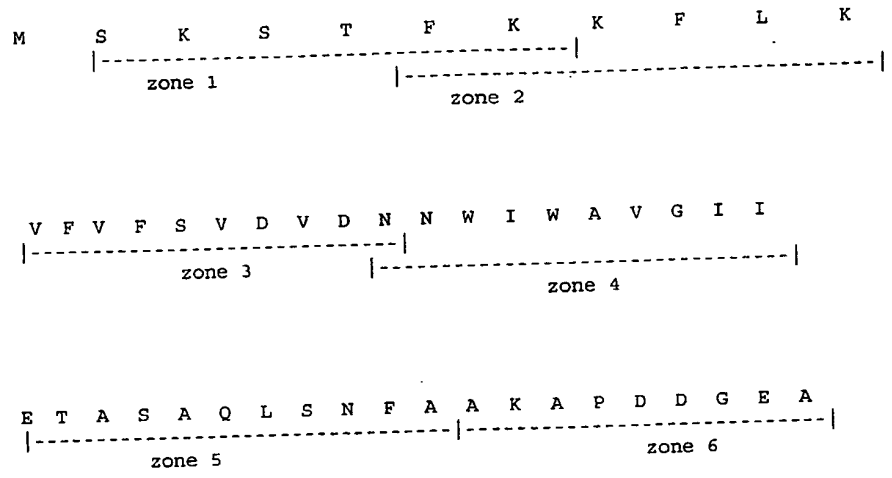
Phage from the library were cycled through rounds of binding selection with an anti-(His)4 antibody (Qiagen) as the capture target. After three or four rounds of selection, individual clones were assayed for hexaHis flag display using a phage ELISA with either the anti-(His)4 antibody or bovine serum albumin (BSA) as target. Of 72 clones assayed, six exhibited at least a two-fold greater signal when captured with the anti-(His)4 antibody rather than with BSA (Fig. 16). These clones were subjected to DNA sequence analysis and the protein sequence was deduced from the DNA sequence.

These protein sequences represent a new class of phage coat proteins which we designate "Protein-12" (P12). The individual unique clones are designated by an additional numeral (*eg.* Protein-12-1 or P12-1). As shown in Figure 16, peptides fused to the C-terminus of P12 are displayed on the surface of M13 phage. The phagemid containing the gene encoding P12-1 was named pS1230a.

Example 25 – Selection of a second generation P12 for the display of a large protein as a Cterminal fusion.

The method of Kunkel (Example 12) was used with the mutagenic oligonucleotide *add-NX* to insert an *Nsi*I restriction site followed by an *Xba*I restriction site into phagemid pS1230a, between the regions encoding P12-1 and the polyHis flag. The resulting sequence was as follows:
...gctgcggctGATGCATCTGGTAGCGTCTAGAGCcaccatcaccatcaccat...(SEQ ID NO: 122)
The inserted sequence is shown in upper case text with the *Nis*I and *Xba*I restriction sites. The inserted sequence is preceded by sequence encoding the final residues of P12-1 and followed by sequence encoding the polyHis flag (both shown in lower case text). The new phagemid was designated pS1232a.

Phagemid pS1232a was digested with *Nsi*I and *Xba*I and a similarly digested DNA fragment encoding an hGH variant (hGH supermutant, hGHsm) with improved afffinity for the hGH binding protein (hGHbp) was inserted. The phagemid was designated pS1239b; it contains an ORF encoding P12-1 followed by a tetrapeptide linker (Ala-Ala-Asp-Ala), followed by hGHsm as shown below. The protein product of the pS1239b ORF is depicted; it consists of P12-1, followed by a tetrapeptide linker (AADA), followed by hGHsm. P12-1 was divided into six zones as indicated, and a library was constructed for each zone. In addition, a linker library was constructed in which random 14-residue peptides were inserted in the middle of the tetrapeptide linker as shown.



(SEQ ID NO. 123)

Phage particles produced from pS1239b did not display hGHsm at levels detectable in a phage ELISA with hGHbp as the capture target.

To obtain a P12 variant capable of displaying hGHsm as a C-terminal fusion, libraries were constructed to vary the sequence of P12-1 encoded by pS1239b. For library construction, a previously described method was used (Example 10). P12-1 was divided into six zones with each zone containing a stretch of contiguous residues (zone 1, residues 2 to 7; zone 2, residues 6 to 11; zone 3, residues 12 to 21; zone 4, residues 21 to 30; zone 5, residues 31 to 40; zone 6, residues 41 to 50). Oligonucleotides were designed to simultaneously replace all codons within the zone with an equal number of degenerate codons (NNS, where N = A, C, G, or T) encoding all twenty natural amino acids. Each oligonucleotide was named according to the zone it mutated (*e.g.* Oligonucleotide *Lib-zone1* mutated zone 1). In addition, an oligonucleotide (*Lib-linker*) was designed to insert 14 degenerate codons (VVC, where V = A, C, or G; encoding Ala, Arg, Asn, Asp, Gly, His, Pro, Ser, or Thr) into the middle of the tetrapeptide linker connecting P12-1 to hGHsm. The diversities of these libraries were as follows: zone 1, 2.5×10^{10}; zone 2, 2.5×10^{10}; zone 3, 2.6×10^{10}; zone 4, 2.4×10^{10}; zone 5, 2.4×10^{10}; zone 6, 2.3×10^{10}; linker library, 2.8×10^{10}.

Phage from all the libraries were pooled and cycled through rounds of binding selection with hGHbp (Example 11) coated on 96-well Maxisorp immunoplates as a target. Phage were propogated in *E. coli* SS320 cells with M13-VCS helper phage (Stratagene). After four rounds of binding selection, individual clones were analyzed for hGHsm display. For each of the rounds 2, 3, and 4, 24 clones were analyzed. Phage were isolated from each clone and hGHsm display was

detected using a phage ELISA (Example 13). A single clone from round 2 exhibited 10-fold greater binding to plates coated with hGHbp in comparison with plates coated with BSA; all other clones exhibited similar binding to either hGHbp or BSA coated plates. The phagemid corresponding to the positive clone was designated pS1258.

The complete DNA sequence of the P12-1 variant encoding ORF of pS1258 was determined and the protein sequence was deduced and shown below. The amino acid numbering is shown to the right.

```
    ATG  AGC  AAG  AGC  ACT  TTC  AAA  AAG  TTT  CTG
    M    S    K    S    T    F    K    K    F    L

    AAA  GTT  TTT  GTT  TTT  TCT  GTT  GAT  GTT  GAT
    K    V    F    V    F    S    V    D    V    D

    AAT  AAT  TGG  ATT  TGG  GCT  GTC  GGT  ATT  ATT
    N    N    W    I    W    A    V    G    I    I

    TAC  ATG  CTC  CTC  GTG  GAG  GCG  TCG  CCC  TGG
    Y    M    L    L    V    E    A    S    P    W

    GCT  GCT  AAG  GCG  CCA  GAC  GAT  GGT  GAA  GCT
    A    A    K    A    P    D    D    G    E    A
```

(SEQ ID NO. 124)

The new variant coat protein was named P12-7; its sequence differs from that of P12-1 in zone 5. The fusion of hGHsm to the C-terminus of P12-7 permits the display of hGHsm on the surface of M13 phage, as evidenced by a phage ELISA.

We also wished to demonstrate that P12-7 permits the display of other proteins, for example wild-type hGH. A phagemid analogous to pS1239b (described above) was constructed and designated pS1239a, with the only difference being that pS1239b encodes a fusion protein consisting of P12-1 followed by wild-type hGH (Example 8). Phage particles produced from pS1239a did not display hGH at levels detectable in a phage ELISA. The method of Kunkel (Example 12) was used with a mutagenic oligonucleotide (*add-P12-7*) to convert the pS1239a DNA sequence encoding P12-1 to DNA sequence encoding P12-7. The new phagemid was designated pW930a; it contains an ORF encoding a fusion protein consisting of P12-7 followed by wild-type hGH. Phage particles isolated from *E. coli* cultures harboring pW930a displayed hGH on their surface, as evidenced by a phage ELISA .

Example 26 – Optimization of linker sequences for display of a peptide fused to the C-terminus of the protein III C-terminal domain.

Oligonucleotides for Example 26:

UH-L4: TTCACCTCGAAAGCAAGCNNSNNSNNSNNSCACCAT
CACCATCACCAT (SEQ ID NO: 125)

UH-L5: TTCACCTCGAAAGCAAGCNNSNNSNNSNNSNNSCACCAT
CACCATCACCAT (SEQ ID NO: 126)

UH-L6: TTCACCTCGAAAGCAAGCNNSNNSNNSNNSNNSNNSCACCAT
CACCATCACCAT (SEQ ID NO: 127)

UH-L8: TTCACCTCGAAAGCAAGCNNSNNSNNSNNSNNSNNSVVCVVCCACCAT
CACCATCACCAT (SEQ ID NO: 128)

UH-L10: TTCACCTCGAAAGCAAGCNNSNNSNNSNNSNNSNNSVVCVVC
VVCVVCCACCATCACCATCACCAT (SEQ ID NO: 129)

UHg3-L6: CTGCGTAATAAGGAGTCTNNSNNSNNSNNSNNSNNS
CACCATCACCATCACCATTAATCATGCCAGTTCTTTTGG (SEQ ID NO: 130)

UHg3-L8: CTGCGTAATAAGGAGTCTNNSNNSNNSNNSNNSNNS
NNSNNSCACCATCACCATCACCATTAATCATGCCAGTTCTTTTGG (SEQ ID NO: 131)

UHg3-L10: CTGCGTAATAAGGAGTCTNNSNNSNNSNNSNNSNNSNNS
NNSNNSNNSCACCATCACCATCACCATTAATCATGCCAGTTCTTTTGG
(SEQ ID NO: 132)

Standard molecular biology techniques were used to construct a phagemid designated pS1428d. Phagemid pS1428d is similar to pS1290a, except that the ORF under the control of the IPTG-inducible Ptac promoter (New England Biolabs) consists of the maltose binding protein signal peptide followed by the C-terminal domain of M13 protein III (Lowman *et al.*, (1991) *Biochemistry*, 30:10832). The method of Kunkel (Example 12) was used to fuse libraries to the C-terminus of the protein III C-terminal domain encoded by pS1428d. The libraries consisted of random linkers of various lengths followed by a hexaHis flag (HHHHHH). The end result was libraries containing ORFs which encoded the C-terminal domain of protein III, followed by random polypeptide linker sequences, followed by the hexaHis flag. The lengths of the linkers were varied and depended on the mutagenic oligonucleotides used: oligonucleotides *UHg3-L6*, *UHg3-L8*, or *UHg3-L10* introduced linkers containing 6, 8, or 10 residues respectively. The diversities of the libraries were as follows: *UHg3-L6*, 3.5×10^{10}; *UHg3-L8*, 1.2×10^{10}; *UHg3-L10*, 2.8×10^{10}.

Phage from the libraries were pooled together and cycled through binding selection with an anti-(His)4 antibody (Qiagen) as the capture target. After two rounds of selection, individual

clones were assayed for hexaHis flag display using a phage ELISA with the anti-(His)4 antibody as target. Three clones exhibiting strong signals were subjected to DNA sequence analysis and the selected linker sequences are shown below.

g3-1	GGG	CAG	GCC	AGG	ATC	GTC	TAC	CGG	CAG	AAG
(SEQ ID NO. 133)										
	G	Q	A	R	I	V	Y	R	Q	K
(SEQ ID NO. 134)										
g3-2	AGG	ATC	AGG	GTC	CTG	CAG	AAG	GGC	AAG	GAG
(SEQ ID NO.135)										
	R	I	R	V	L	Q	K	G	K	E
(SEQ ID NO. 136)										
g3-3	CGC	GCC	AAG	ATC	GAG	CAG	ATC	TGC	AAG	GAG
(SEQ ID NO 137)										
	R	A	K	I	E	Q	I	C	K	E
(SEQ ID NO. 138)										

The sequences shown were inserted between the final residue of the protein III C-terminal domain and a hexaHis flag. For each selectant, the DNA sequence is shown with the deduced amino acid sequence below. A designation for each sequence is shown to the left.

The levels of polyHis flag display were compared with levels of display achieved with C-terminal or N-terminal fusions to protein VIII. Interestingly, display with C-terminal fusion to the protein III C-terminal domain was equivalent to display with N-terminal fusion to protein VIII and about 10-fold higher than display with C-terminal fusion to protein VIII (Fig. 18).

The foregoing written specification is considered to be sufficient to enable one skilled in the art to practice the invention. The present invention is not to be limited in scope by the cultures deposited, since the deposited embodiments are intended as separate illustrations of certain aspects of the invention and any cultures that are functionally equivalent are within the scope of this invention. The deposit of material herein does not constitute an admission that the written description herein contained is inadequate to enable the practice of any aspect of the invention, including the best mode thereof, nor is it to be construed as limiting the scope of the claims to the specific illustrations that it represents. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description and fall within the scope of the appended claims.

While the invention has necessarily been described in conjunction with preferred embodiments, one of ordinary skill, after reading the foregoing specification, will be able to effect various changes, substitutions of equivalents, and alterations to the subject matter set forth herein, without departing from the spirit and scope thereof. Hence, the invention can be practiced in ways other than those specifically described herein. It is therefore intended that the protection granted by Letters Patent hereon be limited only by the appended claims and equivalents thereof.

WHAT IS CLAIMED:

1. A fusion protein comprising a heterologous polypeptide fused to a major coat protein of a virus, wherein the major coat protein is a variant of a wild type major coat protein of the virus.

2. The fusion protein of claim 1, wherein the virus is selected from the group consisting of a filamentous phage, a lambda phage, a Baculovirus, a T4 phage and a T7 phage.

3. The fusion protein of claim 1, wherein the phage is a filamentous phage, the major coat protein is gpVIII and the heterologous polypeptide is fused to the N-terminus or the C-terminus thereof.

4. The fusion protein of claim 1, wherein the major coat protein is a filamentous phage coat protein variant which contains at least one amino acid residue selected from the list below in the position indicated:

Residue Number	Amino Acid Residue
1	E, L, V, Q, D, I, N
2	R, H, F, W, E, K, Y, D
3	T, E, L, S, D, I, V, A
4	D, R, H, E, K
5	R, H, N, D, K, Q, E
6	Y, W, S, I, L, F, T, V
7	T, N, S
8	D, H, R, E, K
9	E, Q, T, D, N, S
11	W, I, V, Y, L, F
12	R, H, N, E, D, K, Q
13	I, L, E, Q, A, V, D, T, N, S
14	L, I, V
15	D, R, N, E, K, H, Q
16	E, V, L, T, D, I, A, S, G
17	E, V, L, I, A, T, D
18	L, V, I
19	L, T, Q, E, I, V, S, A, N, D
20	R, D, H, N, Q, K, E
21	W, Y, I, L, F, V
22	W, F, Y

23	W, Y, I, V, H, K, F, L, R
24	I, Q, L, N, V
25	S, L, I, T, V
26	A, I, V, G, L, M
27	N, T, S
28	I, L, V
29	K, R, F, W, H, Y
30	I, V, L.

5. A fusion protein comprising a heterologous polypeptide fused to at least a portion of a coat protein of a filamentous phage, wherein the coat protein is a variant of a wild type coat protein of the phage, the variant having an alteration in the transmembrane domain or in the cytoplasmic domain of the coat protein.

6. The fusion protein of claim 5, wherein the coat protein is gpIII of a filamentous phage and the heterologous polypeptide is fused to the N-terminus or the C-terminus thereof.

7. The fusion protein of claim 1, wherein the variant has 2 - 50 altered residues relative to the wild type coat protein sequence.

8. The fusion protein of claim 1, wherein the heterologous polypeptide is an antibody or a fragment thereof or a cytokine or a cytokine receptor.

9. A replicable expression vector comprising a gene fusion, wherein the gene fusion encodes the fusion protein of claim 1.

10. A library comprising a plurality of the replicable expression vectors of claim 9, the expression vectors comprising a plurality of different gene fusions encoding a plurality of fusion proteins.

11. Host cells comprising the vector of claim 20.

12. A virus displaying the fusion protein of claim 1 on the surface thereof.

13. A library of virus, comprising a plurality of the virus of claim 12 displaying a plurality of different fusion proteins on the surface thereof.

14. A method, comprising:

constructing a library of phage or phagemid particles displaying a plurality of the fusion protein of claim 1;

contacting the phage or phagemid particles with a target molecule so that at least a portion of the particles bind to the target molecule; and

separating the particles that bind from those that do not bind.

15. A method of decreasing the detection limit of a phage display system utilizing a phage containing a gene fusion encoding a fusion protein, where the gene fusion comprises a first gene encoding a heterologous polypeptide and a second gene encoding at least a portion of a phage coat protein, the method comprising mutating the second gene to encode a variant of a wild type coat protein of the phage.

16. A method of transforming cells, comprising electroporating cells in the presence of heterologous DNA under conditions suitable to allow the heterologous DNA to enter the cells, wherein the heterologous DNA is purified by affinity purification.

17. The method of claim 16, wherein the heterologous DNA is present at a concentration of about 1 picogram to about 500 micrograms/mL.

18. The method of claim 16, obtaining at least 1×10^{10} transformants in one electroporating step.

19. The method of claim 16, wherein the cells are present at a concentration of about 1×10^{11} to about 4×10^{11} cfu/mL.

20. The method of claim 19, wherein the cells are F'::Tn10 *proA*$^{+}$*B*$^{+}$*lacI*^{q}*D(lacZ)M15*/ F-*araD139D(ara-leu)7696galE15galK16D(lac)X74rpsL*(Str^{r})*hsdR2*($r_k^{-}m_k^{+}$)*mcrAmcrB1* *E. coli* cells.

21. A method for producing a product polypeptide, comprising the steps of:

(1) culturing a host cell transformed with a replicable expression vector, the replicable expression vector comprising DNA encoding a product polypeptide operably linked to a control sequence capable of effecting expression of the product polypeptide in the host cell;

wherein the DNA encoding the product polypeptide has been obtained by a method comprising the steps of:

(a) constructing a family of variant replicable plasmids comprising a transcription regulatory element operably linked to a gene fusion encoding a fusion protein, wherein the gene fusion comprises a first gene encoding a polypeptide and a second gene encoding at least a portion of a phage coat protein, wherein the variant replicable plasmids comprise variant first genes encoding variant polypeptides;

(b) transforming suitable host cells with the plasmids using the method of claim 16;

(c) optionally, when the plasmid is a phagemid which requires a helper phage to produce phage particles, infecting the transformed host cells with an amount of helper phage encoding the phage coat protein sufficient to produce recombinant phagemid particles, preferably wherein no more than a minor amount of the phagemid particles display one or more copies of the fusion protein on the surface of the phagemid particles;

(d) culturing the transformed infected host cells under conditions suitable for forming recombinant phage particles containing at least a portion of the plasmid and capable of transforming the host cells;

(e) contacting the recombinant phage particles with a target molecule so that at least a portion of the phage particles bind to the target molecule;

(f) separating phage particles that bind to the target molecule from those that do not bind;

(g) selecting one of the variant polypeptides encoded by the plasmid in a phage particle which binds or does not bind to the target molecule as the product polypeptide and cloning DNA encoding the product polypeptide into the replicable expression vector; and

(2) recovering the expressed product polypeptide.

22. A fusion protein comprising at least a portion of a protein III or protein VIII filamentous phage coat protein having a heterologous polypeptide fused to the carboxyl-terminus thereof.

23. A replicable expression vector comprising a gene fusion, wherein the gene fusion encodes the fusion protein of claim 22.

24. A library comprising a plurality of the replicable expression vectors of claim 23, the expression vectors comprising a plurality of different gene fusions encoding a plurality of fusion proteins.

25. Host cells comprising the vector of claim 23.

26. A virus displaying the fusion protein of claim 22 on the surface thereof.

27. A library of virus, comprising a plurality of the virus of claim 26 displaying a plurality of different fusion proteins on the surface thereof.

28. A method, comprising:

constructing a library of phage or phagemid particles displaying a plurality of the fusion protein of claim 22 on the surface thereof;

contacting the phage or phagemid particles with a target molecule so that at least a portion of the particles bind to the target molecule; and

separating the particles that bind from those that do not bind.

A

Zone 1: codons 1 to 10

	1	2	3	4	5	6	7	8	9	10	
Library	*RNT*	*NAS*	*RNT*	*NAS*	*NAS*	*NYC*	*RNT*	*RNA*	*RNT*	*RNT*	(SEQ ID NO. 139)
	A	D	A	D	D	A	A	A	A	A	
	D	E	D	E	E	F	D	E	D	D	
	G	H	G	H	H	I	G	G	G	G	
	I	K	I	K	K	L	I	I	I	I	
	N	N	N	N	N	P	N	K	N	N	
	S	Q	S	Q	Q	S	S	R	S	S	
	T	Y	T	Y	Y	T	T	T	T	T	
	V		V			V	V	V	V	V	
wt	*GCC*	*GAG*	*GGT*	*GAC*	*GAT*	*CCC*	*GCA*	*AAA*	*GCG*	*GCC*	(SEQ ID NO. 140)
	A	E	G	D	D	P	A	K	A	A	(SEQ ID NO. 141)
1a	*GAT*	*AAG*	*AGT*	*GAG*	*AAG*	*TTC*	*GCT*	*AGA*	*GAT*	*GCT*	(SEQ ID NO. 142)
	D	K	S	E	K	F	A	R	D	A	(SEQ ID NO. 143)
1b	*AAT*	*AAG*	*GAT*	*GAG*	*CAG*	*TTC*	*GCT*	*AGA*	*GCT*	*GCT*	(SEQ ID NO. 144)
	I	K	D	E	Q	F	A	R	A	A	(SEQ ID NO. 145)
1c	*ATT*	*TAC*	*ATT*	*AAG*	*GAG*	*ACC*	*AGT*	*AAA*	*AAT*	*GCT*	(SEQ ID NO. 146)
	I	Y	I	K	E	T	S	K	N	A	(SEQ ID NO. 147)
1d	*AAT*	*TAC*	*GTT*	*GAC*	*CAG*	*GTC*	*AGT*	*AAA*	*AAT*	*GCT*	(SEQ ID NO. 148)
	N	Y	V	D	Q	V	S	K	N	A	(SEQ ID NO. 149)
1e	*GCT*	*AAG*	*GCT*	*GAG*	*GAG*	*TTC*	*GCT*	*GAA*	*GCT*	*GCT*	(SEQ ID NO. 150)
	A	K	A	E	E	F	A	E	A	A	(SEQ ID NO. 151)
1f	*GCT*	*GAC*	*ATT*	*GAC*	*GAC*	*TTC*	*GCT*	*AGA*	*AGT*	*GCT*	(SEQ ID NO. 152)
	A	D	I	D	D	F	A	R	S	A	(SEQ ID NO. 153)

Figure 1A

B

Zone 2: codons 11 to 20

Library	11	12	13	14	15	16	17	18	19	20	
	NWT	*NAS*	*RNT*	*NYT*	*NAS*	*RNT*	*RNT*	*RNT*	*RNT*	*NAS*	(SEQ ID NO. 154)
	D	D	A	A	D	A	A	A	A	D	
	F	E	D	F	E	D	D	D	D	E	
	H	H	G	I	H	G	G	G	G	H	
	I	K	I	L	K	I	I	I	I	K	
	L	N	N	P	N	N	N	N	N	N	
	N	Q	S	S	Q	S	S	S	S	Q	
	V	Y	T	T	Y	T	T	T	T	Y	
	Y		V	V		V	V	V	V		
wt	*TTT*	*AAC*	*TCC*	*CTG*	*CAA*	*GCC*	*TCA*	*GCG*	*ACC*	*GAA*	(SEQ ID NO. 155)
	F	N	S	L	Q	A	S	A	T	E	(SEQ ID NO. 156)
2a	*TAT*	*GAG*	*GCT*	*CTT*	*GAG*	*GAT*	*ATT*	*GCT*	*ACT*	*AAC*	(SEQ ID NO. 157)
	Y	E	A	L	E	D	I	A	T	N	(SEQ ID NO. 158)
2b	*TAT*	*GAG*	*GCT*	*CTT*	*GAG*	*GAT*	*ATT*	*GCT*	*ACT*	*AAC*	(SEQ ID NO. 159)
	Y	E	A	L	E	D	I	A	T	N	(SEQ ID NO. 160)
2c	*TAT*	*GAG*	*GCT*	*CTT*	*GAG*	*GAT*	*ATT*	*GCT*	*ACT*	*AAC*	(SEQ ID NO. 161)
	Y	E	A	L	E	D	I	A	T	N	(SEQ ID NO. 162)
2d	*TAT*	*GAC*	*GTT*	*CTT*	*CAG*	*ATT*	*GCT*	*GCT*	*ATT*	*AAC*	(SEQ ID NO. 163)
	Y	D	V	L	Q	I	A	A	I	N	(SEQ ID NO. 164)
2e*	*CTT*	*AAG*	*GAT*	*CTT*	*AAG*	*GCT*	*ACT*	*GTT*	*ATT*	*CAG*	(SEQ ID NO. 165)
	L	K	D	L	K	A	T	V	I	Q	(SEQ ID NO. 166)
2f*	*TAT*	*GAG*	*ACT*	*ATT*	*AAG*	*GAT*	*GAT*	*ATT*	*GTT*	*AAG*	(SEQ ID NO. 167)
	Y	E	T	I	K	D	D	I	V	K	(SEQ ID NO. 168)
2g*	*CTT*	*CAG*	*AAT*	*ATT*	*CAC*	*AGT*	*AGT*	*ATT*	*AGT*	*AAG*	(SEQ ID NO. 169)
	L	Q	N	I	H	S	S	I	S	K	(SEQ ID NO. 170)
2h*	*TAT*	*AAG*	*ACT*	*GTT*	*CAG*	*GGT*	*GCT*	*ATT*	*GCT*	*AAG*	(SEQ ID NO. 171)
	Y	K	T	V	Q	G	A	I	A	K	(SEQ ID NO. 172)
2i*	*TAT*	*AAG*	*ACT*	*ATT*	*AAG*	*AGT*	*ATT*	*GCT*	*AAT*	*AAG*	(SEQ ID NO. 173)
	Y	K	T	I	K	S	I	A	N	K	(SEQ ID NO. 174)
2j*	*TAT*	*TAG*	*AGT*	*CTT*	*CAG*	*ATT*	*ATT*	*GCT*	*GCT*	*CAG*	(SEQ ID NO. 175)
	Y	Q	S	L	Q	I	I	A	A	Q	(SEQ ID NO. 176)
2k*	*TTT*	*CAG*	*AGT*	*CTT*	*AAG*	*GAT*	*ACT*	*GCT*	*GAT*	*GAG*	(SEQ ID NO. 177)
	F	Q	S	L	K	D	T	A	D	E	(SEQ ID NO. 178)
2m*	*TTT*	*GAG*	*AAT*	*CTT*	*TAG*	*GCT*	*ACT*	*ATT*	*ACT*	*AAG*	(SEQ ID NO. 179)
	F	E	N	L	Q	A	T	I	T	K	(SEQ ID NO. 180)

Figure 1B

C

Zone 3: codons 21 to 30

	21	22	23	24	25	26	27	28	29	30
L	*NWC*	*NWC*	*NKT*	*NWC*	*NYT*	*NKG*	*NYT*	*NKG*	*NWT*	*NWT* (SEQ ID NO. 181)
i	D	D	C	D	A	G	A	G	D	D
b	F	F	F	F	F	L	F	L	F	F
r	H	H	G	H	I	M	I	M	H	H
a	I	I	I	I	L	R	L	R	I	I
r	L	L	L	L	P	V	P	V	L	L
y	N	N	R	N	S	W	S	W	N	N
	V	V	S	V	T		T		V	V
	Y	Y	V	Y	V		V		Y	Y
wt	*TAT*	*ATC*	*GGT*	*TAT*	*GGG*	*TGG*	*GGG*	*ATG*	*GTT*	*GTT* (SEQ ID NO. 182)
	Y	I	G	Y	A	W	A	M	V	V (SEQ ID NO. 183)
3a	*CTT*	*TTC*	*TTT*	*CTC*	*CTT*	*GGG*	*ACT*	*GTG*	*CAT*	*CTT* (SEQ ID NO. 184)
	L	F	F	L	L	G	T	V	H	L (SEQ ID NO. 185)
3b	*TAC*	*TAC*	*CTT*	*AAC*	*ATT*	*TTG*	*GCT*	*GTG*	*TAT*	*GTT* (SEQ ID NO. 186)
	Y	Y	L	N	I	L	A	V	Y	V (SEQ ID NO. 187)
3c	*TTC*	*ATC*	*CGT*	*GTC*	*ACT*	*TGG*	*ACT*	*ATG*	*TAT*	*GTT* (SEQ ID NO. 188)
	F	I	R	V	T	W	T	M	Y	V (SEQ ID NO 189)
3d	*GTC*	*ATC*	*CGT*	*TAC*	*GTT*	*ATG*	*TCT*	*ATG*	*TAT*	*GTT* (SEQ ID NO. 190)
	V	I	R	Y	V	M	S	M	Y	V (SEQ ID NO. 191)

Figure 1C

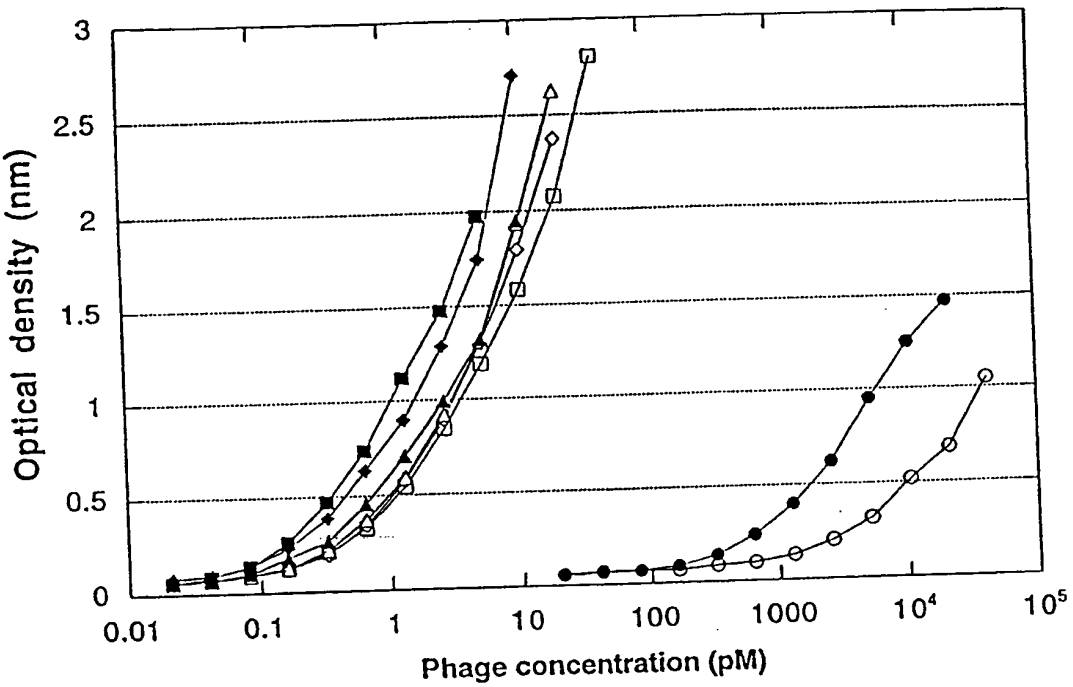

Optical density (nm)
3
2.5
2
1.5
1
0.5
0
0.01
0.1
1
10
100
1000
10^4
10^5
Phage concentration (pM)

Figure 2

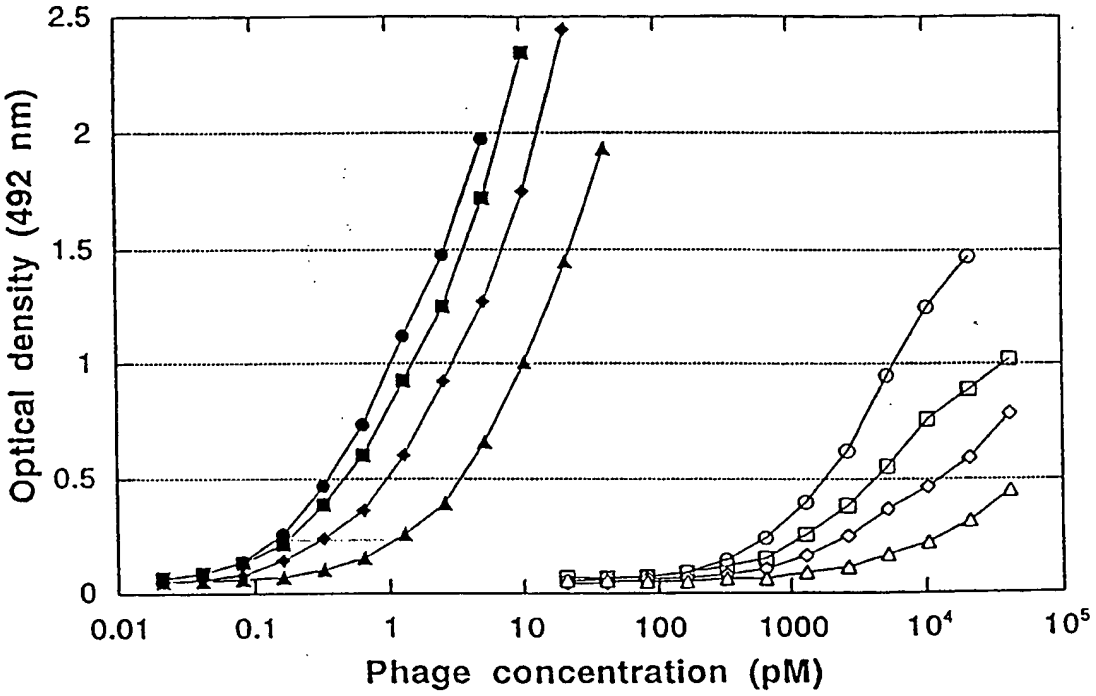

Optical density (492 nm)
2.5
2
1.5
1
0.5
0
0.01
0.1
1
10
100
1000
10⁴
10⁵
Phage concentration (pM)

Figure 3

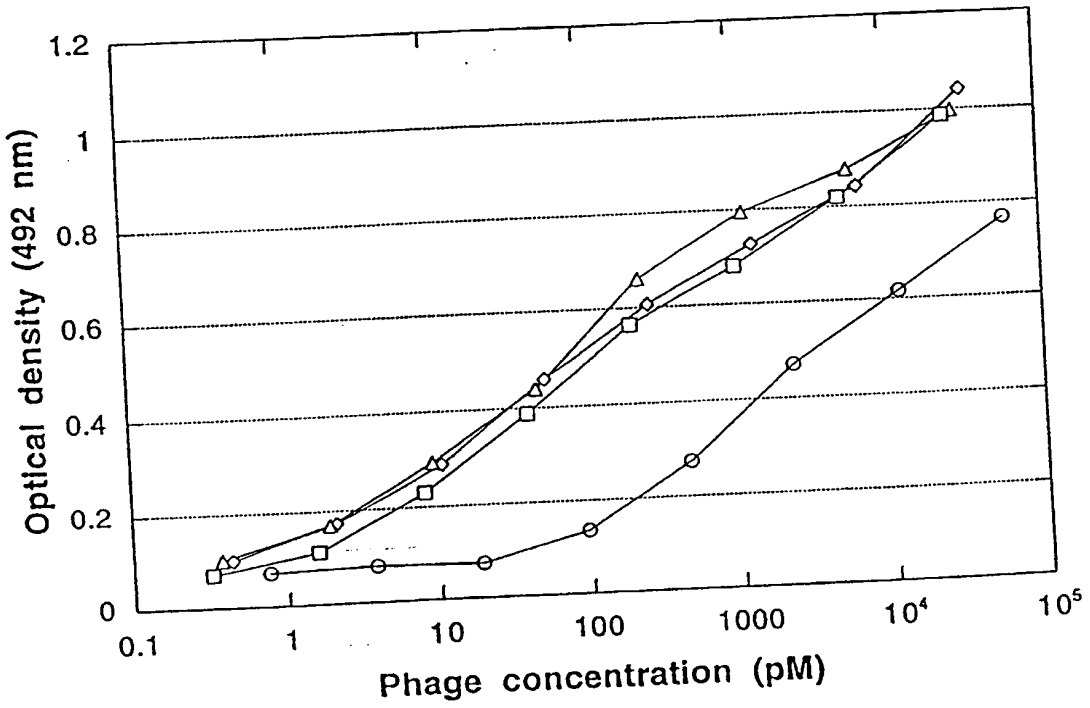

1.2
1
0.8
0.6
0.4
0.2
0
Optical density (492 nm)
0.1
1
10
100
1000
10⁴
10⁵
Phage concentration (pM)

Figure 4A

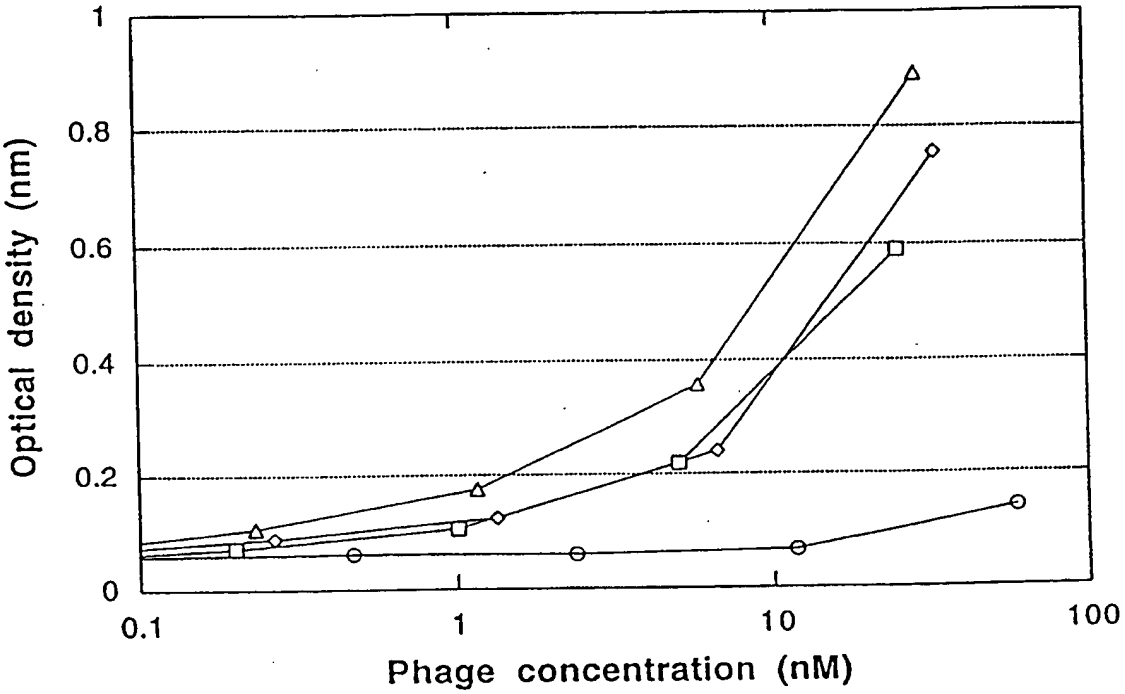

1
0.8
0.6
0.4
0.2
0
Optical density (nm)
0.1
1
10
100
Phage concentration (nM)

Figure 4B

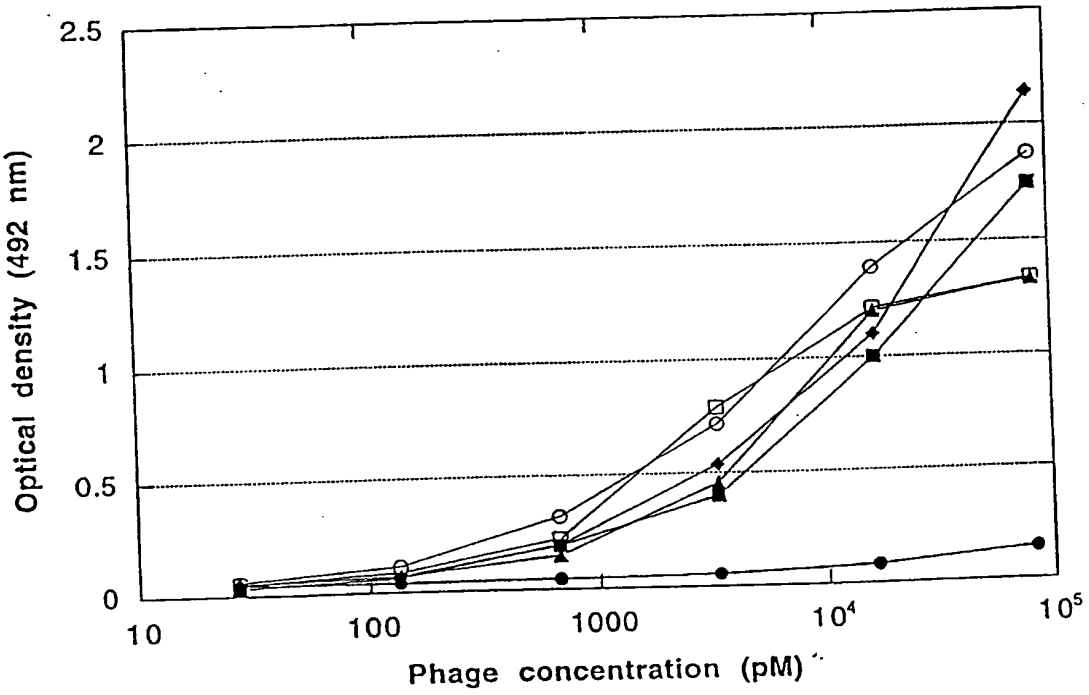

2.5
2
1.5
1
0.5
0
Optical density (492 nm)
10
100
1000
10^4
10^5
Phage concentration (pM)

Figure 5

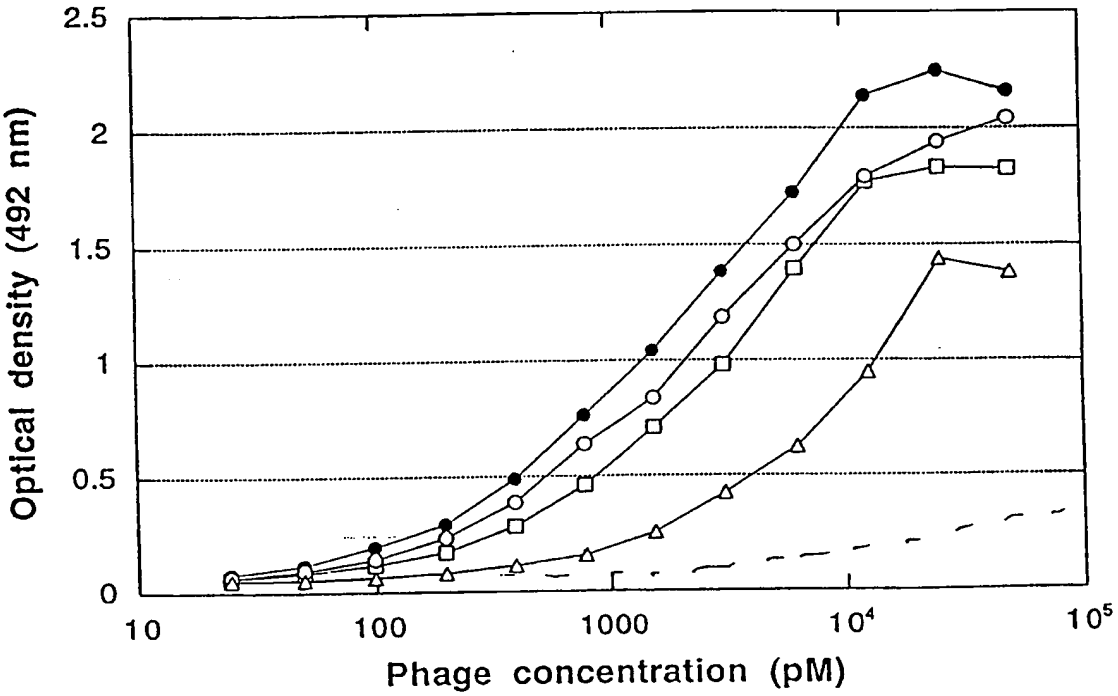

2.5
2
1.5
1
0.5
0
Optical density (492 nm)
10
100
1000
10⁴
10⁵
Phage concentration (pM)

Figure 6

A

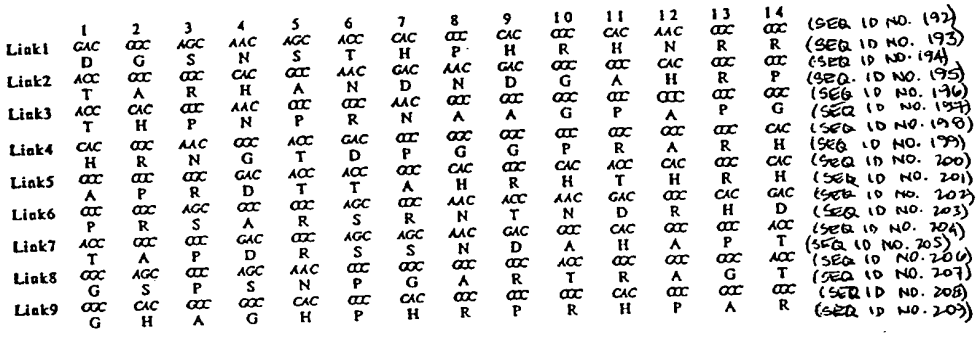

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Link1 GAC D | ACC G | AGC S | AAC N | AGC S | ACC T | CAC H | CCC P | CAC H | CCC R | CAC H | AAC N | CCC R | CCC R (SEQ ID NO. 192)
Link2 ACC T | CCC A | CCC R | CAC H | CCC A | AAC N | GAC D | AAC N | GAC D | CCC G | CCC A | CAC H | CCC R | CCC P (SEQ ID NO. 193)
Link3 ACC T | CAC H | CCC P | AAC N | CCC P | CCC R | AAC N | CCC A | CCC D | CCC A | CCC G | CCC P | CCC R | CCC G (SEQ ID NO. 194)
Link4 CAC H | CCC R | AAC N | CCC G | ACC T | GAC D | CCC P | CCC G | CCC G | CCC P | CCC R | CCC A | CCC R | CAC H (SEQ ID NO. 195)
Link5 CCC A | CCC P | CCC R | GAC D | ACC T | ACC T | CCC A | CAC H | CCC R | CAC H | ACC T | CCC H | CAC R | CAC H (SEQ ID NO. 196)
Link6 CCC P | CCC R | AGC S | CCC A | CCC R | AGC S | CCC R | AAC N | ACC T | GAC D | GAC D | CCC R | CAC H | GAC D (SEQ ID NO. 197)
Link7 ACC T | CCC A | CCC P | GAC D | CCC R | AGC S | AGC S | AAC N | GAC D | CCC A | CAC H | CCC A | CCC P | ACC T (SEQ ID NO. 198)
Link8 CCC G | AGC S | CCC P | AGC S | AAC N | CCC P | CCC G | CCC A | CCC R | ACC T | CCC R | CAC H | CCC A | ACC T (SEQ ID NO. 199)
Link9 CCC G | CAC H | CCC A | CCC G | CAC H | CCC P | CAC H | CCC R | CCC P | CCC R | CAC H | CCC P | CCC A | CCC R (SEQ ID NO. 200)

Figure 7A

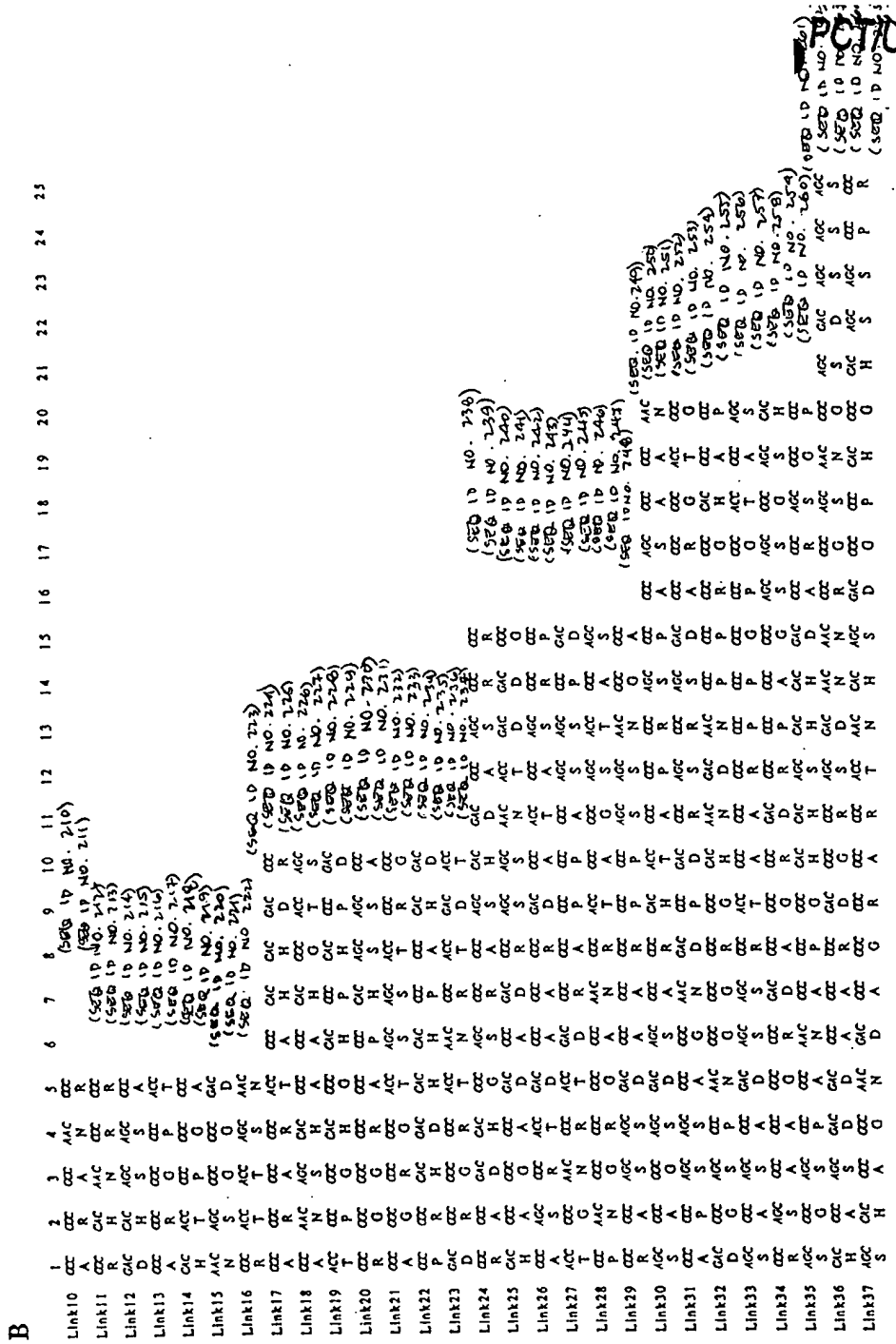


Figure 7B

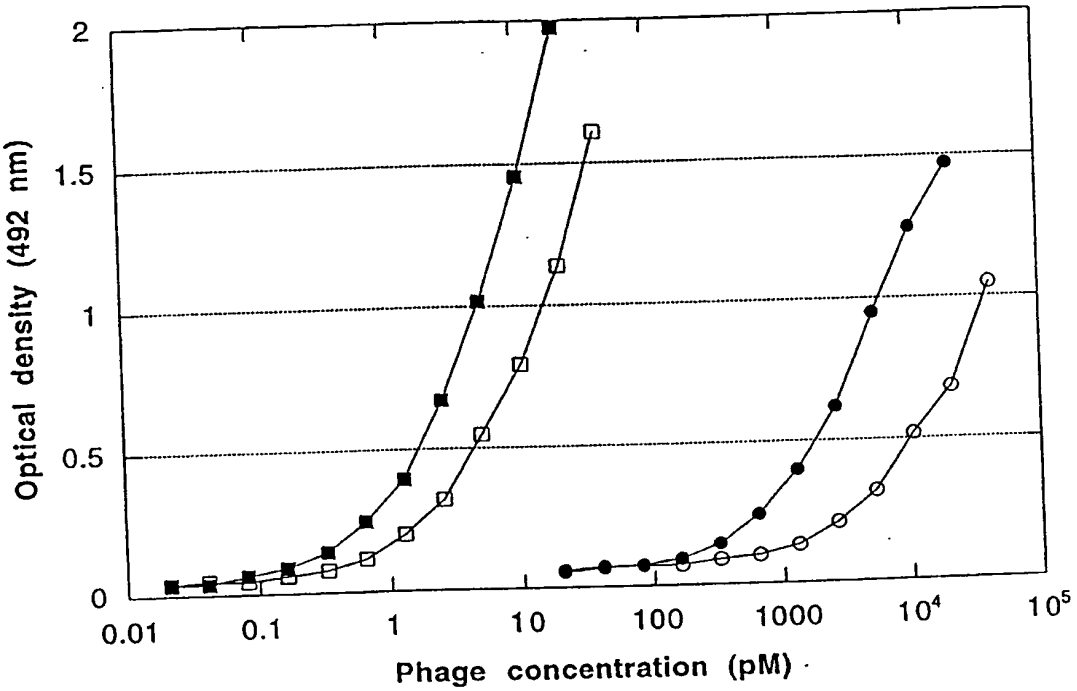

Optical density (492 nm)
2
1.5
1
0.5
0
0.01
0.1
1
10
100
1000
10^4
10^5
Phage concentration (pM)

Figure 8A

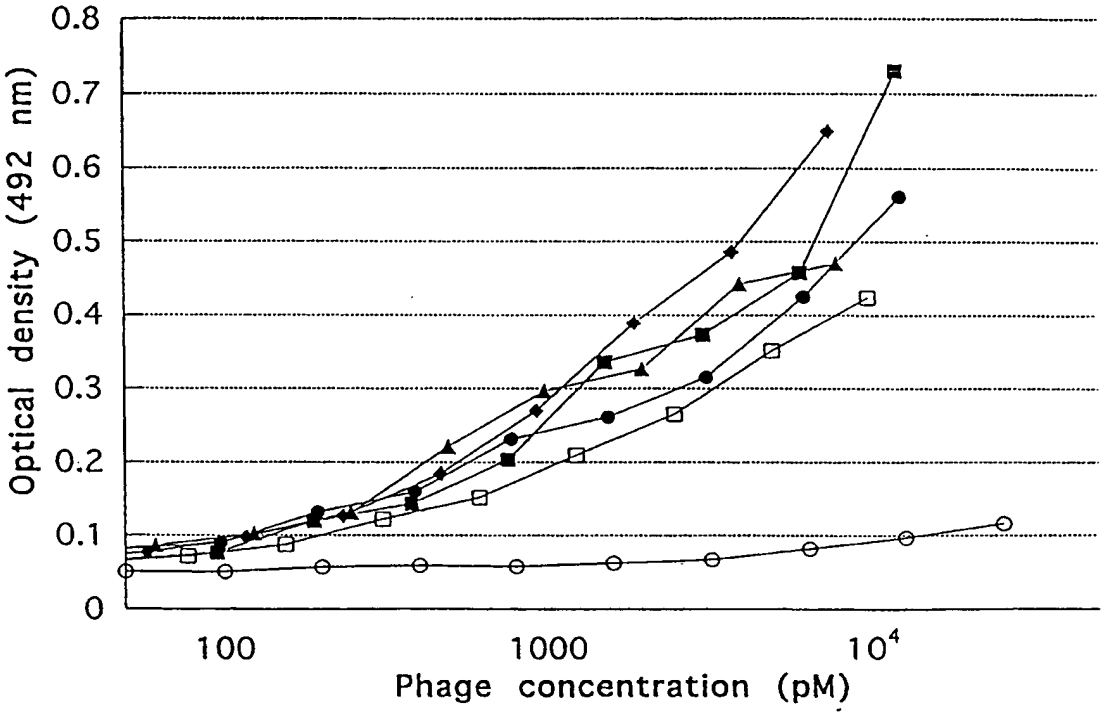

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Optical density (492 nm)
100
1000
10^4
Phage concentration (pM)

Figure 8B

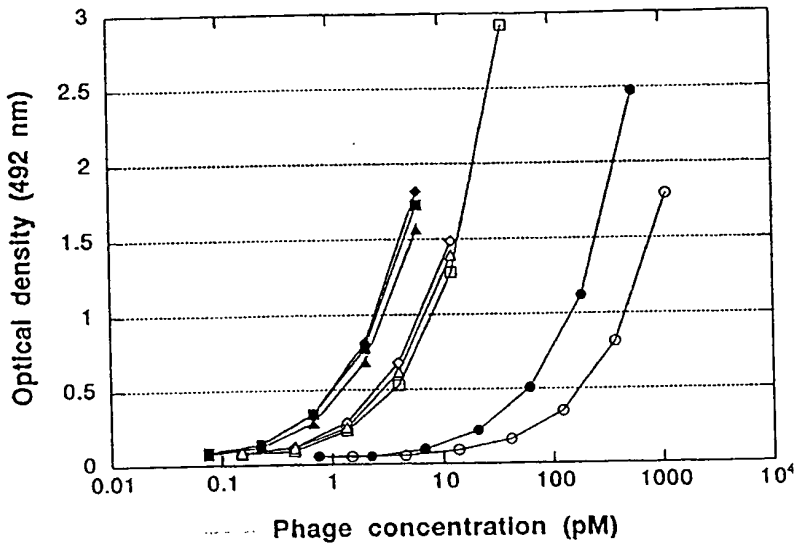

Optical density (492 nm)
3
2.5
2
1.5
1
0.5
0
0.01
0.1
1
10
100
1000
10⁴
Phage concentration (pM)

Figure 9

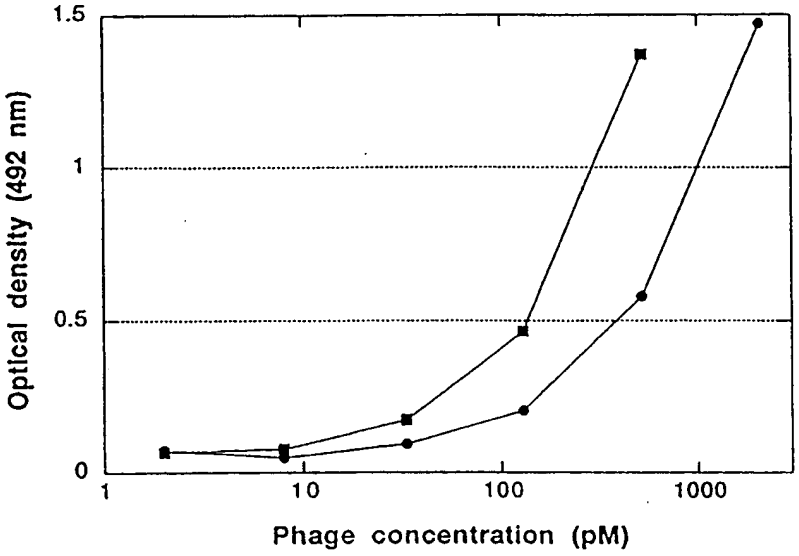

1.5
1
0.5
0
Optical density (492 nm)
1
10
100
1000
Phage concentration (pM)

Figure 10

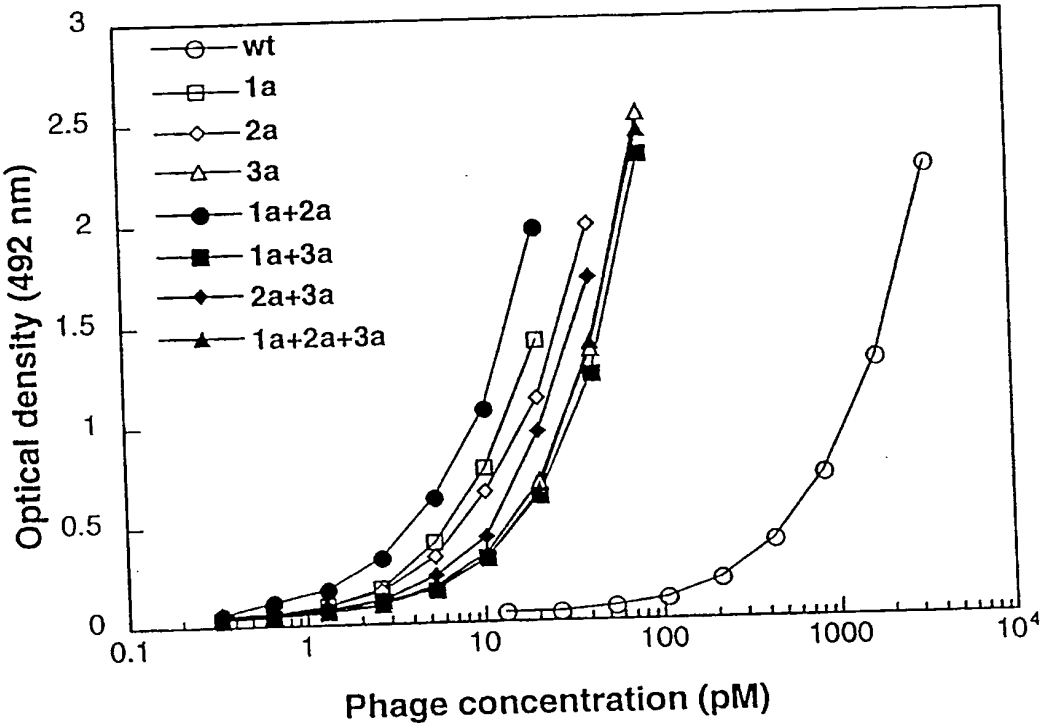

wt
1a
2a
3a
1a+2a
1a+3a
2a+3a
1a+2a+3a
Optical density (492 nm)
Phage concentration (pM)
0
0.5
1
1.5
2
2.5
3
0.1
1
10
100
1000
10⁴

Figure 11

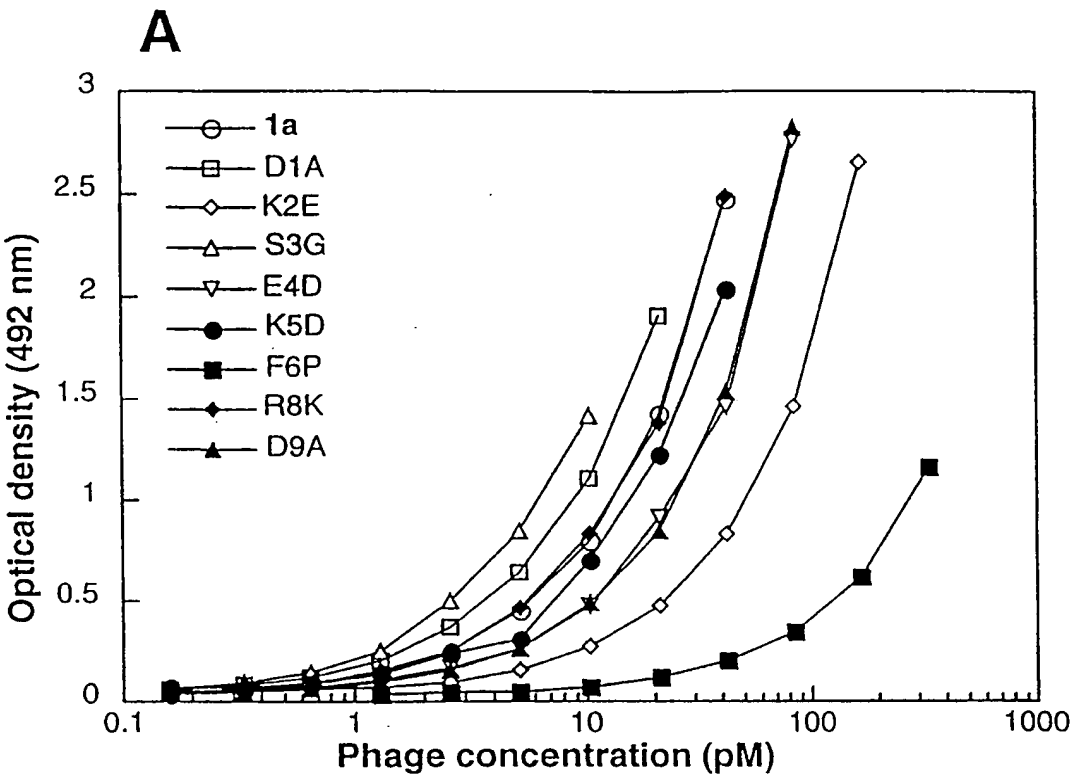

A
Optical density (492 nm)
Phage concentration (pM)
1a
D1A
K2E
S3G
E4D
K5D
F6P
R8K
D9A
0
0.5
1
1.5
2
2.5
3
0.1
1
10
100
1000

Figure 12A

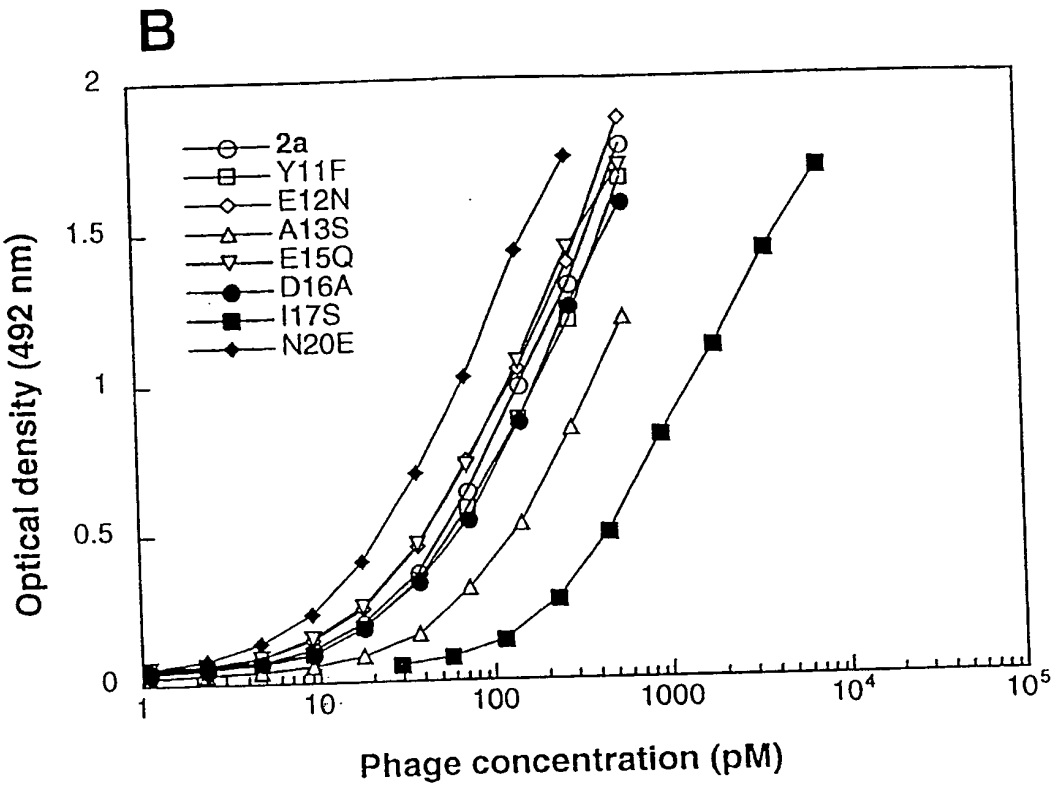

B
2
1.5
1
0.5
0
Optical density (492 nm)
2a
Y11F
E12N
A13S
E15Q
D16A
I17S
N20E
1
10
100
1000
10^4
10^5
Phage concentration (pM)

Figure 12B

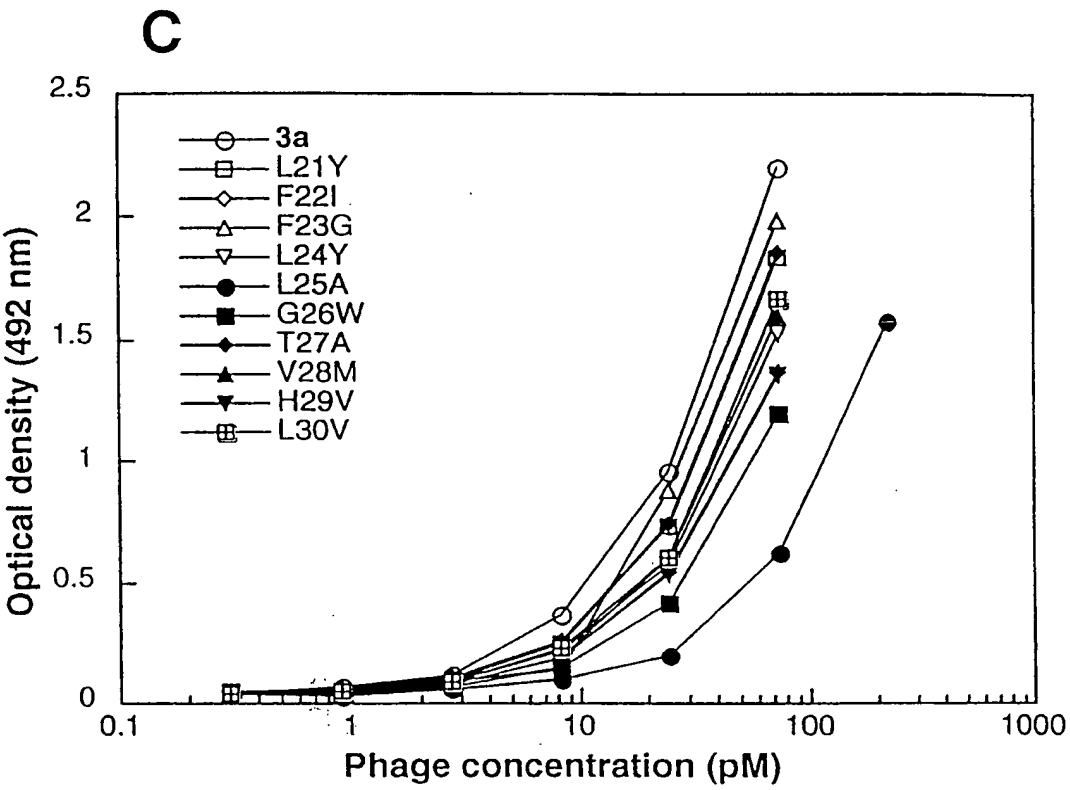

C
2.5
2
1.5
1
0.5
0
Optical density (492 nm)
3a
L21Y
F22I
F23G
L24Y
L25A
G26W
T27A
V28M
H29V
L30V
0.1
1
10
100
1000
Phage concentration (pM)

Figure 12C

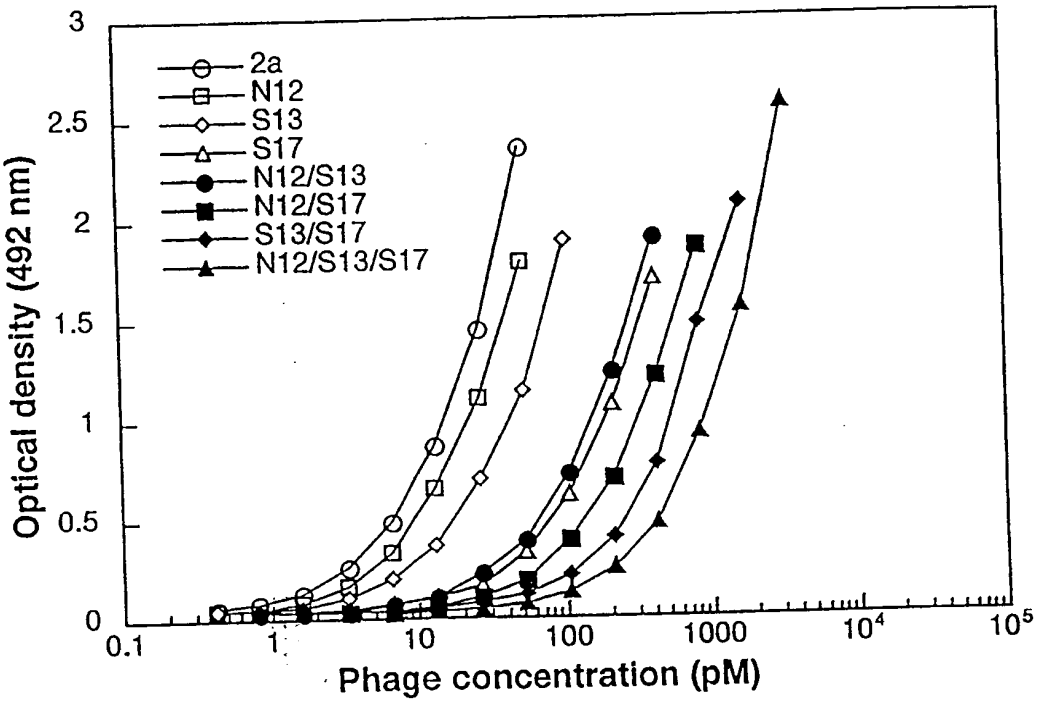

Optical density (492 nm)
Phage concentration (pM)
2a
N12
S13
S17
N12/S13
N12/S17
S13/S17
N12/S13/S17
0
0.5
1
1.5
2
2.5
3
0.1
1
10
100
1000
10^4
10^5

Figure 12D

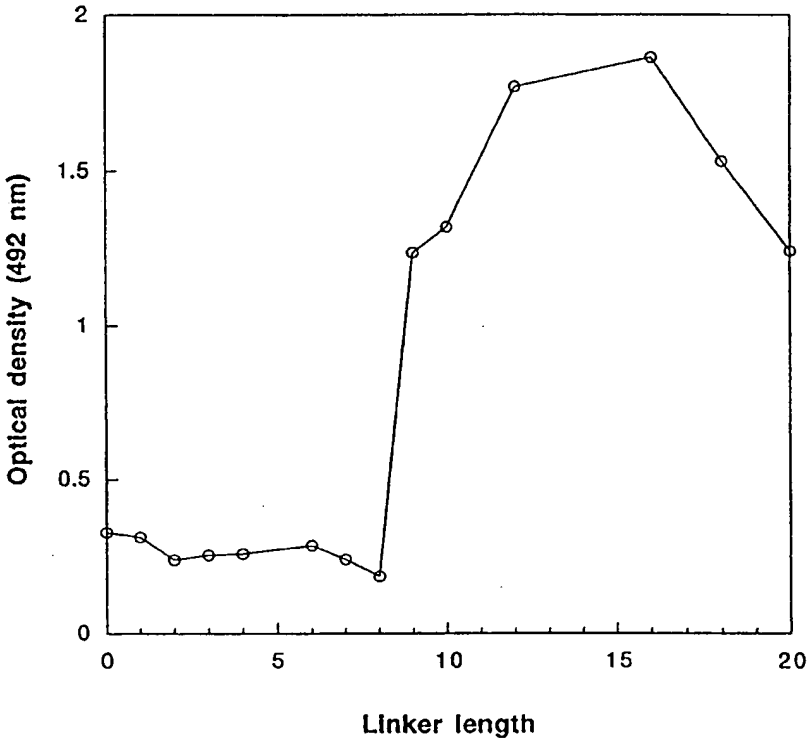


Figure 13

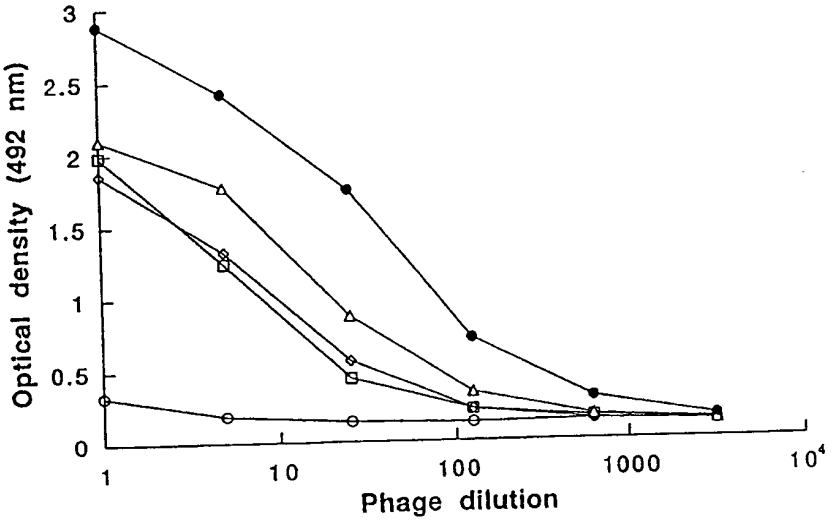

Optical density (492 nm)
3
2.5
2
1.5
1
0.5
0
1
10
100
1000
10⁴
Phage dilution

Figure 14

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
	NWT	NKT	NWT	NYT	NYT	NKT	NWT	NWT	NWT	NWT	NWT	NKG	NYT	NKG	NYT	NWC	NKT	NWT	NWT	(SEQ ID NO. 266).
L	F	F	F	F	F	F	F	F	F	F	F	L	F	L	F	F	F	F	F	
I	L	L	L	L	L	L	L	L	L	L	L	M	L	M	L	L	L	L	L	
B	I	I	I	I	I	I	I	I	I	I	I	V	I	V	I	I	I	I	I	
R	V	V	V	V	V	V	V	V	V	V	V	W	V	W	V	V	V	V	V	
A	N	C	N	S	S	C	N	N	N	N	N	R	S	R	S	Y	C	N	N	
R	H	R	H	P	P	R	H	H	H	H	H	G	P	G	P	H	R	H	H	
Y	N	S	N	T	T	S	N	N	N	N	N		T		T	N	S	N	N	
	D	G	D	A	A	G	D	D	D	D	D		A		A	D	G	D	D	
P12-1	GTT	TTT	GTT	TTT	TCT	GTT	GAT	GTT	GAT	AAT	AAT	TGG	ATT	TGG	GCT	GTC	GGT	ATT	GTT	(SEQ ID NO. 267)
	V	F	V	F	S	V	D	V	D	N	N	W	I	W	A	V	G	I	V	(SEQ ID NO. 268)
P12-2	CAT	AGT	CTT	GCT	GTT	ATT	GAT	GAT	AAT	TTT	TAT	TGG	GTT	GGG	TTT	TAC	GGT	TAT	GTT	(SEQ ID NO. 269)
	H	S	L	A	V	I	D	D	N	F	Y	W	V	G	F	Y	G	Y	V	(SEQ ID NO. 270)
P12-3	CTT	TTT	TAT	CCT	GTT	AGT	GTT	CAT	ATT	GTT	ATT	CGG	TTT	TTG	TCT	CTC	TTT	CTT	GTT	(SEQ ID NO. 271)
	L	F	Y	P	V	S	V	H	I	V	I	R	F	L	S	L	F	L	V	(SEQ ID NO. 272)
P12-4	CTT	AGT	GTT	GTT	GTT	CGT	GAT	CTT	ATT	TAT	AAT	GTG	GTT	ATG	TTT	CAC	GTT	GTT	AAT	(SEQ ID NO. 273)
	L	S	V	V	V	R	D	L	I	Y	N	V	V	M	F	H	V	V	N	(SEQ ID NO. 274)
P12-5	CTT	GGT	TTT	TCT	ACT	CGT	GTT	CTT	GTT	GAT	GAT	TGG	CTT	ATG	GTT	AAC	AGT	TTT	GAT	(SEQ ID NO. 275)
	L	G	F	S	T	R	V	L	V	D	D	W	L	M	V	N	S	F	D	(SEQ ID NO. 276)
P12-6	TAT	TTT	CTT	GCT	TTT	AGT	ATT	ATT	GAT	CTT	TTT	AGG	CTT	TGG	CTT	TAC	TTT	GTT	AAT	(SEQ ID NO. 277)
	Y	F	L	A	F	S	I	I	D	L	F	R	L	W	L	Y	F	V	N	(SEQ ID NO. 278)

Figure 15

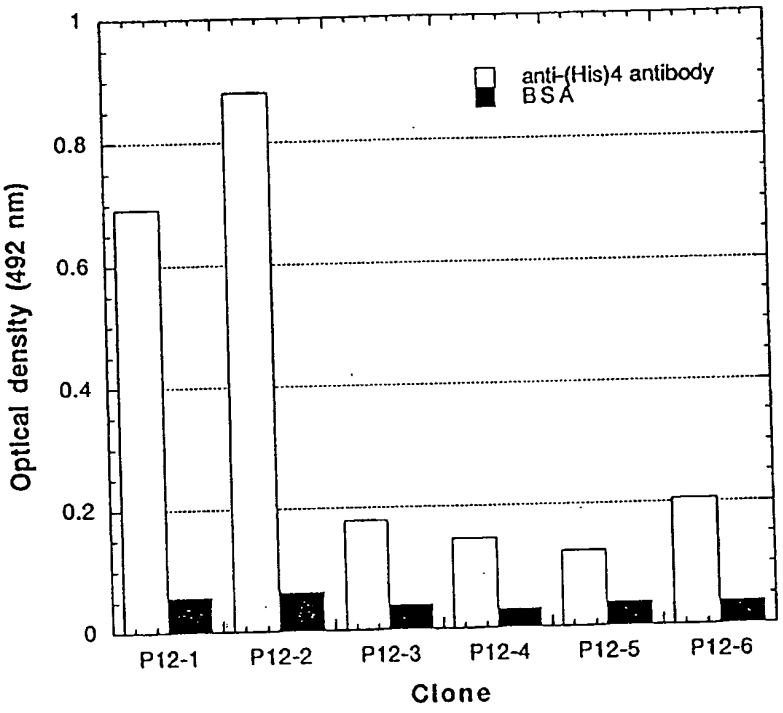

1
0.8
0.6
0.4
0.2
0
Optical density (492 nm)
anti-(His)4 antibody
BSA
P12-1
P12-2
P12-3
P12-4
P12-5
P12-6
Clone

Figure 16

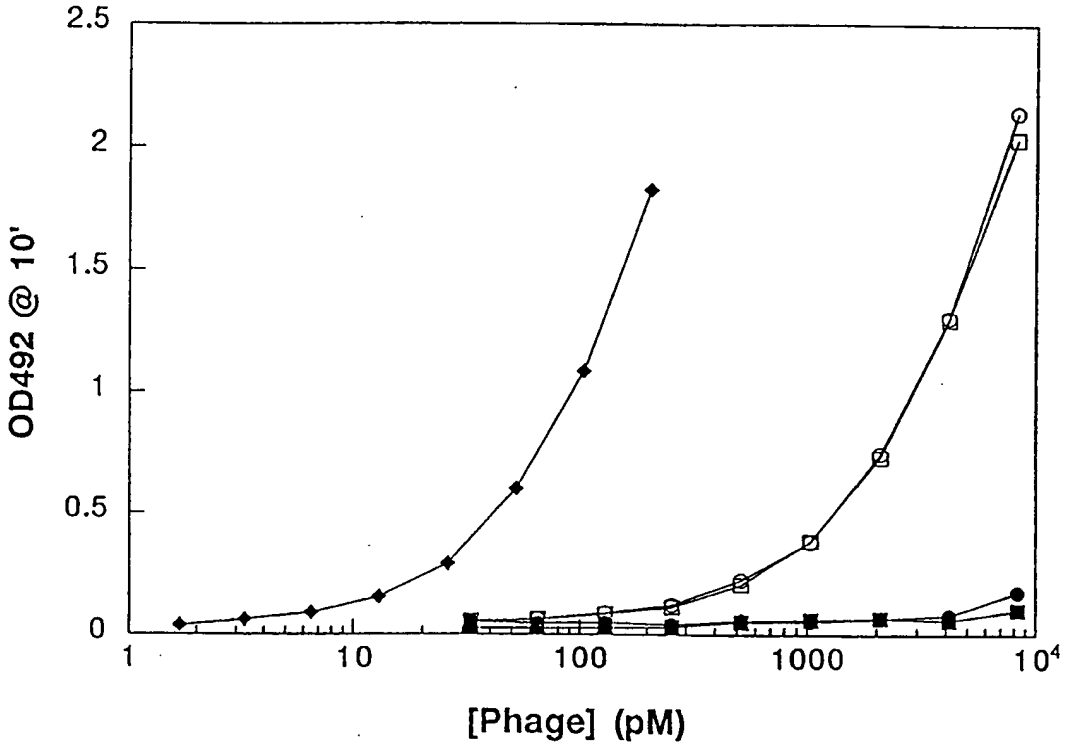


Figure 17

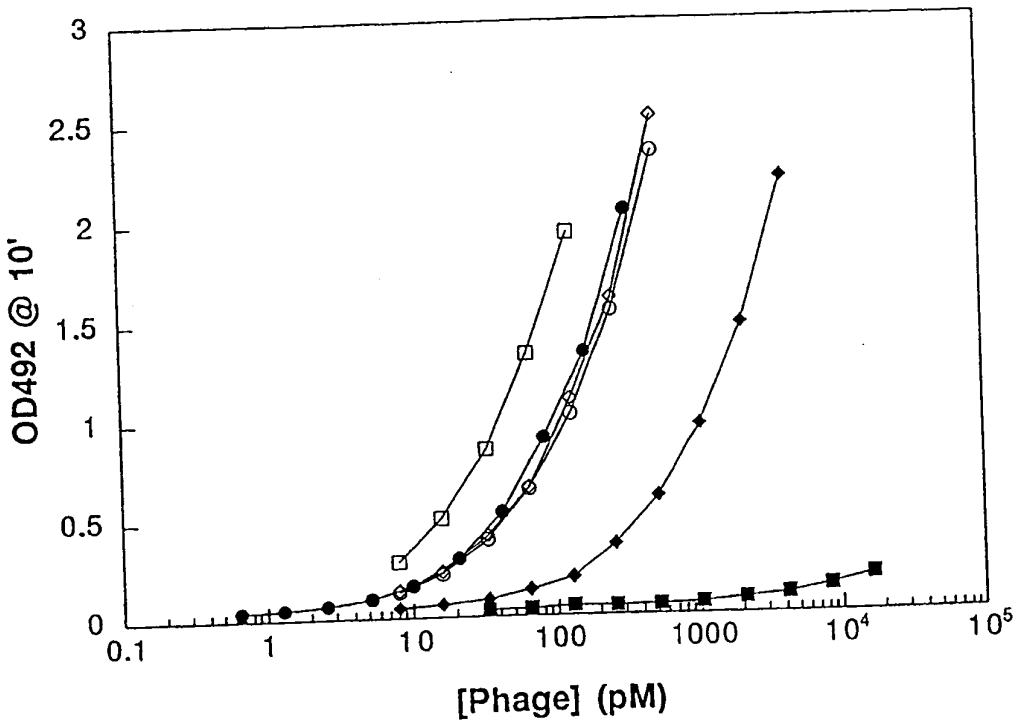


Figure 18

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

PCT WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 7 :		(11) Internationale Veröffentlichungsnummer: WO 00/46376
C12N 15/36, C07K 14/02, C12N 15/62, C07K 19/00, C12N 15/87	A2	(43) Internationales Veröffentlichungsdatum: 10. August 2000 (10.08.00)

(21) Internationales Aktenzeichen: PCT/DE00/00363

(22) Internationales Anmeldedatum: 4. Februar 2000 (04.02.00)

(30) Prioritätsdaten:
199 04 800.2 5. Februar 1999 (05.02.99) DE

(71)(72) Anmelder und Erfinder: HILDT, Eberhard [DE/DE]; Robert Koch Institut, Nordufer 20, D–13353 Berlin (DE).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: HOFSCHNEIDER, Peter [DE/DE]; Nördliche Auffahrtsallee 65, D–80638 München (DE).

(74) Anwalt: HUBER, Bernard; Huber & Schüssler, Truderinger Strasse 246, D–81825 München (DE).

(81) Bestimmungsstaaten: US, europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht
Ohne internationalen Recherchenbericht und erneut zu veröffentlichen nach Erhalt des Berichts.

(54) Title: PARTICLES FOR GENE THERAPY

(54) Bezeichnung: PARTIKEL ZUR GENTHERAPIE

(57) Abstract

The invention relates to particles comprising: a) a protein membrane with a fusion protein which comprises a virus protein, a cell–permeability–mediating peptide and a heterologous cell–specific binding site; and b) a nucleic acid which is contained in the protein membrane and presents sequences for a virus–specific packaging signal and a structural gene. The invention also relates to methods for producing such particles, means suitable for this purpose and the use of the particles in gene therapy.

(57) Zusammenfassung

Die vorliegende Erfindung betrifft Partikel, umfassend: (a) eine Proteinhülle mit einem Fusionsprotein, das ein Virus–Protein, ein Zellpermeabilität–vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle umfasst, und (b) eine in der Proteinhülle vorliegende Nukleinsäure, die Sequenzen für ein Virus–spezifisches Verpackungssignal und ein Struktur–Gen aufweist. Ferner betrifft die Erfindung Verfahren zur Herstellung solcher Partikel und hierfür geeignete Mittel sowie die Verwendung der Partikel zur Gentherapie.

LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien
AM	Armenien
AT	Österreich
AU	Australien
AZ	Aserbaidschan
BA	Bosnien-Herzegowina
BB	Barbados
BE	Belgien
BF	Burkina Faso
BG	Bulgarien
BJ	Benin
BR	Brasilien
BY	Belarus
CA	Kanada
CF	Zentralafrikanische Republik
CG	Kongo
CH	Schweiz
CI	Côte d'Ivoire
CM	Kamerun
CN	China
CU	Kuba
CZ	Tschechische Republik
DE	Deutschland
DK	Dänemark
EE	Estland
ES	Spanien
FI	Finnland
FR	Frankreich
GA	Gabun
GB	Vereinigtes Königreich
GE	Georgien
GH	Ghana
GN	Guinea
GR	Griechenland
HU	Ungarn
IE	Irland
IL	Israel
IS	Island
IT	Italien
JP	Japan
KE	Kenia
KG	Kirgisistan
KP	Demokratische Volksrepublik Korea
KR	Republik Korea
KZ	Kasachstan
LC	St. Lucia
LI	Liechtenstein
LK	Sri Lanka
LR	Liberia
LS	Lesotho
LT	Litauen
LU	Luxemburg
LV	Lettland
MC	Monaco
MD	Republik Moldau
MG	Madagaskar
MK	Die ehemalige jugoslawische Republik Mazedonien
ML	Mali
MN	Mongolei
MR	Mauretanien
MW	Malawi
MX	Mexiko
NE	Niger
NL	Niederlande
NO	Norwegen
NZ	Neuseeland
PL	Polen
PT	Portugal
RO	Rumänien
RU	Russische Föderation
SD	Sudan
SE	Schweden
SG	Singapur
SI	Slowenien
SK	Slowakei
SN	Senegal
SZ	Swasiland
TD	Tschad
TG	Togo
TJ	Tadschikistan
TM	Turkmenistan
TR	Türkei
TT	Trinidad und Tobago
UA	Ukraine
UG	Uganda
US	Vereinigte Staaten von Amerika
UZ	Usbekistan
VN	Vietnam
YU	Jugoslawien
ZW	Zimbabwe

Partikel zur Gentherapie

Die vorliegende Erfindung betrifft Nukleinsäure enthaltende Partikel, die gezielt an Zellen binden und in diese ihre Nukleinsäure einführen können. Ferner betrifft die Erfindung Verfahren zur Herstellung solcher Partikel und hierfür geeignete Mittel sowie die Verwendung der Partikel zur Gentherapie.

Für eine Gentherapie ist es wichtig ein Gentransfersystem zu haben, das spezifisch ist, d.h. mit dem gewünschte Zellen erreicht und in diese Gene eingeführt werden können. Bei Leberzellen kann dies prinzipiell mit einem modifizierten Hepatitis B-Virus (HBV) als Vektor möglich, da HBV leberzellspezifisch ist. Für andere Zellen, z.B. Fibroblasten, existiert jedoch kein Gentransfersystem, das befriedigende Ergebnisse liefert.

Der vorliegenden Erfindung liegt somit die Aufgabe zugrunde, ein Gentransfersystem bereitzustellen, das spezifisch ist, d.h. mit dem gewünschte Zellen erreicht und in diese Gene eingeführt werden können.

Erfindungsgemäß wird dies durch die Gegenstände in den Patentansprüchen erreicht.

Die vorliegende Erfindung beruht auf den Erkenntnissen des Anmelders, daß Nukleinsäure enthaltende Partikel, die ein Fusionsprotein aufweisen, welches ein Virus-Protein, ein Zellpermeabilität-vermittelndes Peptid, insbesondere ein solches der deutschen Patentanmeldung 198 50 718.6, und eine heterologe zellspezifische Bindungsstelle umfaßt, an entsprechende Zellen binden und in diese ihre Nukleinsäure einführen können. Beispielsweise hat er Nukleinsäure enthaltende HBV-Partikel hergestellt, die an Fibroblasten binden und in diese ihre Nukleinsäure einführen. Hierzu hat er

die Hepatocyten-Bindungsstelle, die in der Region PreS1, insbesondere zwischen den Aminosäuren 21-47, des großen Oberflächenproteins von HBV (LHBs) vorliegt, gegen die α5ß1-Integrin-Bindungsstelle von Fibronektin ausgetauscht, wobei das Zellpermeabilität-vermittelnde Peptid, das in der Region PreS2 von LHBs vorliegt, erhalten blieb. Ferner hat er Partikel mit Fibroblasten-Spezifität hergestellt, indem er das Core-Protein von HBV (HBcAg) mit der α5ß1-Integrin-Bindungsstelle von Fibronektin und dem vorstehenden Zellpermeabilität-vermittelnden Peptid verbunden hat. Desweiteren hat er erkannt, daß die in den Partikeln enthaltene Nukleinsäure in den Zellen exprimiert wird.

Erfindungsgemäß werden die Erkenntnisse des Anmelders genutzt, Partikel bereitzustellen, umfassend:

(a) eine Proteinhülle mit einem Fusionsprotein, das ein Virus-Protein, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle umfaßt, und

(b) eine in der Proteinhülle vorliegende Nukleinsäure, die Sequenzen für ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist.

Der Ausdruck "Zellpermeabilität-vermittelndes Peptid" umfaßt jegliche Peptide, die Substanzen, insbesondere Proteinen, eine Zellpermeabilität vermitteln können. Dies sind insbesondere jene Peptide, die in der deutschen Patentanmeldung 198 50 718.6 des Anmelders genannt sind. Besonders bevorzugt ist ein Peptid, das die folgende Aminosäure-(DNA)-Sequenz umfaßt:

P	L	S	S	I	F	S	R	I	G	D	P
CCC	ATA	TCG	TCA	ATC	TTC	TCG	AGG	ATT	GGG	GAC	CCT

Der Ausdruck "zellspezifische Bindungsstelle" umfaßt jegliche Bindungsstellen von Proteinen und sonstigen kleinen Molekülen, über welche die Proteine bzw. die Moleküle an Zellen binden

können. Beispiele solcher Bindungsstellen finden sich in Cytokinen und Wachstumsfaktoren. Ferner finden sie sich in Liganden von Hormon-, Neurotransmitter-, Blutzellenoberflächen- und Integrin-Rezeptoren. Eine bevorzugte Bindungsstelle ist die α5ß1-Integrin-Bindungsstelle von Fibronektin. Diese wird nachstehend mit RGD bezeichnet und umfaßt die Aminosäuren Arginin, Glycin und Aspartat.

Der Ausdruck "Virus" umfaßt DNA- und RNA-Viren, insbesondere Adenoviren, Adeno-assoziierte Viren, Vacciniaviren, Baculoviren, Hepatitis C-Viren, Hepatitis A-Viren, Influenzaviren und Hepadna-Viren. Beispiele letzterer sind HBV, WHV ("woodchuck hepatitis virus") GSHV ("ground squirrel hepatitis virus"), RBSHV ("red-bellied squirrel hepatitis virus") DHV ("Pekin duck hepatitis virus") und HHV ("heron hepatitis virus"), wobei HBV bevorzugt ist.

Der Ausdruck "Virus-Protein" betrifft jegliches Protein eines vorstehenden Virus, das als ganzes oder teilweise in einem Fusionsprotein zusammen mit einem Zellpermeabilität-vermittelnden Peptid und einer heterologen zellspezifischen Bindungsstelle in Form eines weiteren Peptids vorliegen kann. Das Protein kann auch bereits das Zellpermeabilität-vermittelnde Peptid enthalten. Ein Beispiel für ein solches Protein ist LHBs. Dieses wird wie andere Oberflächen-Proteine und Core-Proteine, z.B. HBcAg, bevorzugt. Der Ausdruck "heterolog" weist darauf hin, daß das Protein von Natur aus nicht die vorstehende zellspezifische Bindungsstelle aufweist. Günstig kann es sein, wenn die homologe, d.h. von Natur aus vorliegende Bindungsstelle des Proteins ausgeschaltet ist. Besonders günstig kann es sein, wenn die homologe durch die heterologe Bindungsstelle ersetzt ist.

Der Ausdruck "Nukleinsäure" umfaßt RNA und DNA, wobei beide einzelsträngig und/oder doppelsträngig sein können.

Der Ausdruck "Virus-spezifisches Verpackungssignal" weist auf eine Signalsequenz in vorstehender Nukleinsäure hin, mittels

dieser die Nukleinsäure in die Proteinhülle eines Partikels verpackt wird. Die Signalsequenz ist spezifisch für ein vorstehendes Virus. Eine bevorzugte Signalsequenz ist jene von HBV. Diese findet sich auf der HBV-DNA und wird in der Literatur mit Epsilon bezeichnet.

Der Ausdruck "Struktur-Gen" umfaßt Gene, die für Polypeptide (Proteine) kodieren. Beispiele von Polypeptiden sind Tumornekrosefaktor, Interferone, Interleukine, Lymphokine, Wachstumsfaktoren, Plasmaproteine, z.B. Gerinnungsfaktoren und Stoffwechselenzyme, und Rezeptoren. Insbesondere können die Polypeptide solche sein, welche die Immunogenität von Zellen steigern können. Dies können Polypeptide sein, die Tumorzellen fehlen, z.B. Cytokine, wie IL-2 und GM-CSF, und kostimulatorische Moleküle, wie B7-1, Tumor-assoziierte Antigene, z.B. MAGE1, Tyrosinasen und virale Polypeptide, z.B. E7 von humanem Papillomvirus und EBNA-3-Polypeptid von Epstein-Barr-Virus. Ferner können die Polypeptide Adapter-Polypeptide, Oligomerisierungsmotive eines Polypeptids, Polypeptidfragmente von Virus-Hüllpolypeptiden und Hormone sein. Ferner umfaßt der Ausdruck "Struktur-Gen" Antisense-Oligonukleotide, Peptid-Nukleinsäuren, Consensus-Sequenzen für Transkriptionsfaktoren und Ribozyme.

Erfindungsgemäß werden Partikel bevorzugt, die ein Fusionsprotein enthalten, das ein LHBs oder Fragmente davon und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt. Günstig ist es, wenn die heterologe Bindungsstelle, insbesondere RGD, anstelle der homologen Bindungsstelle vorliegt. Besonders bevorzugt ist es, wenn das Fusionsprotein die Aminosäuresequenz von Fig. 1 oder eine hiervon durch eine oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz aufweist.

Des weiteren werden Partikel bevorzugt, die ein Fusionsprotein enthalten, das ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid, z.B. ein solches der deutschen Patentanmeldung 198 50 718.6, insbesondere mit der vorstehend angegebenen Aminosäu-

resequenz, und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt. Besonders bevorzugt ist es, wenn das Fusionsprotein die Aminosäuresequenz von Fig. 2 oder eine hiervon durch eine oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz aufweist.

Der Ausdruck "eine durch eine oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz" weist darauf hin, daß diese Aminosäuresequenz ein Fusionsprotein kennzeichnet, das vergleichbare Elemente und Funktionen wie das Fusionsprotein von Fig. 1 oder Fig. 2 aufweist, sich allerdings bis zu 20 %, vorzugsweise 10 %, von der Aminosäuresequenz von Fig. 1 oder 2 unterscheidet.

Ein erfindungsgemäßes Partikel kann durch übliche Verfahren hergestellt werden. Enthält das Partikel z.B. ein Fusionsprotein, das ein LHBs umfaßt, in dem die homologe durch eine heterologe Bindungsstelle, insbesondere RGD, ersetzt ist, so ist ein Verfahren günstig, das folgende Verfahrensschritte aufweist:

(a) Co-Transfektion von Zellen, die für ein Heptatitis B-Virus-Genom kodieren, wobei diese kein LHBs exprimieren, mit einem ersten Expressionsvektor, der für ein Fusionsprotein kodiert, das ein LHBs umfaßt, in dem die homologe durch eine heterologe Bindungsstelle, insbesondere RGD, ersetzt ist, und mit einem zweiten Expressionsvektor, der ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist, und

(b) Isolierung und Reinigung des Partikels.

Enthält das Partikel ein Fusionsprotein, das ein HBcAG, ein Zellpermeabilität-vermittelndes Peptid gemäß der deutschen Patentanmeldung 198 50718.6, insbesondere jenes mit vorstehender Aminosäuresequenz, und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt, so ist ein Verfahren

günstig, das folgende Verfahrensschritte umfaßt:

(a) Co-Transfektion von Zellen, die für eine HBV-Polymerase kodieren, mit einem ersten Expressionsvektor, der für ein Fusionsprotein kodiert, das ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid gemäß der deutschen Patentanmeldung 198 50718.6, insbesondere jenes mit vorstehender Aminosäuresequenz, und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt, und mit einem zweiten Expressionsvektor, der ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist, und

(b) Isolierung und Reinigung des Partikels.

Hinsichtlich der Ausdrücke "Expressionsvektor", "Zellen", und "Isolierung und Reinigung" wird auf nachstehende Ausführungen, insbesondere in den Beispielen, verwiesen. Die Zellen stellen ebenfalls einen Gegenstand der vorliegenden Erfindung dar. Bezüglich der anderen Ausdrücke wird auf vorstehende Ausführungen verwiesen.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Fusionsprotein, das ein LHBs oder Fragmente davon und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt. Vorzugsweise umfaßt das Fusionsprotein die Aminosäuresequenz von Fig. 1 oder eine hiervon durch ein oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Fusionsprotein, das ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe Bindungsstelle, insbesondere RGD, umfaßt. Vorzugsweise umfaßt das Fusionsprotein die Aminosäuresequenz von Fig. 2 oder eine hiervon durch eine oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz.

Hinsichtlich des Ausdrucks "eine durch eine oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz" wird auf vorstehende Ausführungen verwiesen.

Ein weiterer Gegenstand der vorliegenden Erfindung ist eine Nukleinsäure, die für ein vorstehendes Fusionsprotein kodiert. Die Nukleinsäure kann eine RNA oder eine DNA sein. Bevorzugt ist eine DNA, die folgendes umfaßt:

(a) Die DNA von Fig. 1 oder 2 oder eine hiervon durch ein oder mehrere Basenpaare unterschiedliche DNA, oder

(b) eine mit der DNA von (a) über den degenerierten genetischen Code verwandte DNA.

Der Ausdruck "eine durch ein oder mehrere Basenpaare unterschiedliche DNA"
weist darauf hin, daß diese DNA für ein Fusionsprotein kodiert, das vorgleichbare Elemente und Funktionen wie das Fusionsprotein von Fig. 1 oder 2 aufweist, sich allerdings von der Basensequenz von Fig. 1 oder 2 derart unterscheidet, daß in der Aminosäuresequenz ein Unterschied von maximal 20 %, vorzugsweise 10 %, vorliegt.

Eine erfindungsgemäße DNA kann als solche oder in einem Vektor vorliegen. Insbesondere kann eine erfindungsgemäße DNA in einem Expressionsvektor vorliegen. Beispiele solcher sind dem Fachmann bekannt. Im Falle eines Expressionsvektors für E. coli sind dies z.B. pGEMEX, pUC-Derivate, pGEX-2T, pET3b und pQE-8. Für die Expression in Hefe sind z.B. pY100 und Ycpad1 zu nennen, während für die Expression in tierischen Zellen z.B. pKCR, pEFBOS, cDM8, pCEV4, pCDNA3, pKSV10, pRCMV und pRK5 anzugeben sind. Für die Expression in Insektenzellen eignet sich besonders der Bacculovirus-Expressionsvektor pAcSGHisNT-A.

Der Fachmann kennt geeignete Zellen, um die erfindungsgemäße, in einem Expressionsvektor vorliegende DNA zu exprimieren. Beispiele solcher Zellen umfassen die E.coli-Stämme HB101,

DH1, x1776, JM101, JM 109, BL21, SG 13009 und M15pRep4, den Hefe-Stamm Saccharomyces cerevisiae, die tierischen Zellen L, NIH 3T3, FM3A, CHO, COS, Vero, HeLa, Hep62, CCL13 und 293, die Insektenzellen Sf9 und Sf21 und die Pflanzenzellen Lupinus albus.

Der Fachmann kennt Verfahren und Bedingungen Zellen mit einem, die erfindungsgemäße DNA enthaltenden Expressionsvektor zu transformieren bzw. transfizieren und die Zellen zu kultivieren. Auch sind ihm Verfahren bekannt, das durch die erfindungsgemäße DNA exprimierte Virus-Protein zu isolieren und zu reinigen.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein gegen ein vorstehendes Fusionsprotein gerichteter Antikörper. Ein solcher Antikörper kann durch übliche Verfahren hergestellt werden. Er kann polyklonal bzw. monoklonal sein. Zu seiner Herstellung ist es günstig, Tiere, insbesondere Kaninchen oder Hühner für einen polyklonalen und Mäuse für einen monoklonalen Antikörper, mit dem Fusionsprotein zu immunisieren. Weitere "Booster" der Tiere können ebenfalls mit dem Fusionsprotein erfolgen. Der polyklonale Antikörper kann dann aus dem Serum bzw. Eigelb der Tiere erhalten werden. Für den monoklonalen Antikörper werden Milzzellen der Tiere mit Myelomzellen fusioniert.

Ein weiterer Gegenstand der vorliegenden Erfindung ist ein Kit. Ein solcher umfaßt eine oder mehrere der folgenden Komponenten:

(a) ein erfindungsgemäßes Fusionsprotein,
(b) eine erfindungsgemäße DNA,
(c) einen erfindungsgemäßen Antikörper, sowie
(d) übliche Hilfsstoffe, wie Träger, Puffer, Lösungsmittel, Kontrollen, etc.

Von den einzelnen Komponenten können jeweils ein oder mehrere Vertreter vorliegen. Hinsichtlich der einzelnen Ausdrücke wird

auf vorstehende Ausführungen verwiesen.

Die vorliegende Erfindung stellt ein Gentransfer-System bereit, das spezifisch ist, d.h. mit dem gewünschte Zellen erreicht und in diese Gene eingeführt werden können. Die Zellen können einzeln oder in einem Gewebe vorliegen. Ferner können die Zellen isoliert oder im Körper eines Individuums vorliegen. Somit eignet sich die vorliegende Erfindung für eine ex vivo bzw. in vivo Gentherapie von Zellen bzw. Geweben. Die Anwendung der vorliegenden Erfindung kann dabei durch erfindungsgemäße Antikörper überwacht und gesteuert werden.

Somit stellt die vorliegende Erfindung einen großen Schritt dar gentherapeutische Veränderungen an Zellen bzw. Geweben gezielt durchzuführen.

Kurze Beschreibung der Zeichnungen.

Fig. 1 zeigt die Aminosäure- und DNA-Sequenzen eines erfindungsgemäßen Fusionsproteins, das ein LHBs und die heterologe Bindungsstelle RGD umfaßt, wobei diese die homolge ersetzt.

Fig. 2 zeigt die Aminosäure- und DNA-Sequenzen eines erfindungsgemäßen Fusionsproteins, das ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid der vorstehenden Aminosäuresequenz und die heterologe Bindungsstelle RGD umfaßt.

Die vorliegende Erfindung wird durch die nachstehenden Beispiele erläutert.

Beispiel 1: Herstellung eines erfindungsgemäßen Partikels, das ein Fusionsprotein enthält, welches ein LHBs und eine heterologe Bindungsstelle umfaßt.

(A) Herstellung eines Expressionsvektor, der für sämtliche HBV-spezifischen Proteine mit Ausnahme von LHBs kodiert.

Hierzu wird von dem Plasmid pTKTHBV2 (vgl. Will et al., Proc. Natl. Acad. Sci. 82 (1985), 891-895) ausgegangen. Dieses enthält zwei Kopien des HBV-Genoms. In einer ersten PCR wird ein Fragment von ntHBV2821 (erste Kopie) bis ntHBV2870 (zweite Kopie) amplifiziert. Der forward Primer (nt 2821-2855) weist folgende Sequenz auf: CCA TAT TCT TGG GAA CAA GAT ATC CAG CAC GGG GC. Eine EcoRV-Schnittstelle ist unterstrichen. Das Triplett ACG zwischen nt 2849-2852 ersetzt das ATG-Startcodon von LHBs. Der backward Primer (nt 2877-2845) weist folgende Sequenz auf: GGA TTG CTG GTG GAA GAT ATC TGC CCC GTG CTG. Eine EcoRV-Schnittstelle ist unterstrichen. Das Triplett CGT zwischen nt 2852-2849 ersetzt das natürliche Triplett CAT. Erhaltene PCR-Fragmente werden mit EcoRV verdaut und über ein präparatives 1%iges Agarosegel gereinigt. Ein Fragment der Größe von ca. 3,3 kb wird aus dem Gel eluiert und aufbewahrt.

In einer zweiten PCR werden ein forward Primer, der eine EcoRV-Schnittstelle gefolgt von der nachstehenden Sequenz ntHBV2860 (zweite Kopie)-2878 (erste Kopie) (CAG CAC GGG GCA GAT ATC TTC CAC CAG CAA TCC) aufweist, und ein backward Primer verwendet, der eine EcoRV Schnittstelle gefolgt von der nachstehenden Sequenz ntHBV 2830-2810 (GC CCC GTG CTG GAT ATC ATC TTG TTC CCA AGA ATA TGG) aufweist. Erhaltende PCR-Fragmente werden mit EcoRV verdaut und über ein präparatives 1 % Agarose-Gel gereinigt. Ein Fragment der erwarteten Größe wird aus dem Gel eluiert und dephosphoryliert. Dieses Fragment wird mit dem vorstehenden ca. 3.3 kb großen in eine Ligasereaktion eingesetzt, wodurch der HBV-Expressionsvektor pTKTHBV2Ldef erhalten wird. Dieser Expressionsvektor kodiert für sämtliche HBV-spezifischen Proteine mit Ausnahme von LHBs.

(B) Herstellung eines Expressionsvektors, der für ein Fusionsprotein kodiert, welches ein LHBs und die heterologe Bindungsstelle RGD umfaßt.

Ausgehend von dem Plasmid pTKTHBV2 (vgl. vorstehend) wird das Fragment ntHBV2990-834 durch PCR amplifiziert. Der 5'-Primer weist folgende Sequenz auf: AAA AGA TCT GGC CGT GGC GAA GGA GCT GGA GCA TTC. Diese umfaßt eine BgIII-Schnittstelle, gefolgt von einem ATG-Startcodon und der für das Tripeptid RGD-kodierenden Sequenz. Der PreS1-spezifische Leserahmen wird genutzt. Der 3'-Primer wiest die folgende Sequenz auf: AAA AGA TCT GGT TTA AAT GTA TAC CCA AAG. Diese umfaßt eine BgIII-Schnittstelle. Erhaltene PCR-Fragmente werden mit BgIII verdaut und in den mit BgIII-gespaltenen und dephosphorylierten Vektor pCDNA.3 (Invitrogen) inseriert, wodurch der Expressionsvektor pCRGDLHBs erhalten wird. Dieser Expressionsvektor kodiert für ein N-terminal verkürztes LHBs, das die RGD-Bindungsstelle umfaßt.

(C) Herstellung eines ein Struktur-Gen und ein Verpackungssignal aufweisenden Expressionsvektors

Es wird eine für das HBV-Verpackungssignal Epsilon kodierende Sequenz, z.B. ntHBV 1840-1914, mittels PCR amplifiziert. Durch die verwendeten Primer wird eine EcoRV-Schnittstelle eingeführt. Die Sequenz des forward Primers lautet: CCC GAT ATC ATG TCA TCT CTT GTT CAT GTC CTA. Die Sequenz des backward Primers lautet: GGG GAT ATC GGT CGA TGT CCA TGC CCC AAA. Erhaltene PCR-Fragmente werden mit EcoRV gespalten und in den mit EcoRV gespaltenen und dephosphorylierten Vektor pCDNA.3 (vgl. vorstehend) inseriert, wodurch der Vektor pcVPHBV erhalten wird. Dieser Vektor enthält das HBV-spezifische Verpackungssignal Epsilon.

Ausgehend von dem Vektor pCeGFP (Invitrogen), der für ein "green fluorescent protein" unter der Kontrolle des CMV-Promotors kodiert, wird die den CMV-Promotor und das GFP-Gen

enthaltende Sequenz mittels PCR amplifiziert. Der forward Primer hat folgende Sequenz: GGG GGA TCC CGA TGT ACG GGC CAG ATA TAC GCG TTG. Der backward Primer hat folgende Sequenz: GGG GGA TCC GCG GCC GCT TTA CTT GTA. Die verwendeten Primer enthalten jeweils eine BamHI-Schnittstelle. Erhaltene PCR-Fragmente werden mit BamHI gespalten und in den mit BamHI gespaltenen und dephosphorylierten Vektor pCVPHBV (Invitrogen) inseriert, wodurch der Expressionsvektor pCVPHBVeGFP erhalten wird. Dieser Expressionsvektor enthält das HBV-spezifische Verpackungssignal Epsilon, den CMV-Promotor und eine für eGFP kodierende Sequenz.

(D) Herstellung einer Verpackungszellinie

Etwa $0.8x10^6$ HepG2-Zellen werden mit 4µg von pTKTHBV2Ldef (vgl. (A)) und 2 µg pCDNA.3 (vgl. (B)) mittels Lipofektion transfiziert. pCDNA.3 kodiert für G418 Resistenz. 2h nach Transfektion werden die Zellen in ein 700 mg G418/l enthaltendes Medium überführt. G418 resistende Klone werden nach 14d subkultiviert. Die stabile Integration von pTKTHBV2Ldef wird mittels PCR und Southern Blots nachgewiesen. Die Expression des Oberflächenproteins SHBs von HBV und von HBcAg wird mittels spezifischer Antikörper in ELISAS nachgewiesen. Es wird die Verpackungszellinie HepG2-TKTHBV2Ldef erhalten. Diese Zellinie exprimiert sämtliche HBV-spezifischen Proteine mit Ausnahme von LHBs.

(E) Herstellung erfindungsgemäßer Partikel

Etwa $0.8x10^6$ Zellen der Verpackungszellinie von (D) werden mit 3µg von pCRGDLHBs (vgl. (B)) und 3µg von pCVPHBVeGFP (vgl. (B)) mittels Lipofektion transfiziert. 72 h nach Transfektion weren die Zellen bzw. deren Überstände gesammelt und einer PEG-Fällung unterworfen. Anschließend wird eine CsCl-Dichtegradienten-Zentrifugation durchgeführt. Es werden erfindungsgemäße Partikel in reiner Form erhalten. Diese Partikel umfassen sämtliche HBV-spezifischen Proteine mit Ausnahme von LHBs, das durch ein RGD-LHBs ersetzt ist.

Beispiel 2: Herstellung eines erfindungsgemäßen Partikels, das ein Fusionsprotein enthält, welches ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe Bindungsstelle umfaßt.

Es wird eine für ein Zellpermeabilität-vermittelndes Peptid (nachstehend mit ZPP bezeichnet) kodierende DNA verwendet. Diese hat folgende Sequenz: XXX AGA TCT ATG CCC ATA TCG TCA ATC TTC TCG AGG ATT GGG GAC CCT GGA TCC XXX (X bezeichnet ein beliebiges Nukleotid). Diese weist am 5'-Ende eine BglII-Schnittstelle, gefolgt von einem ATG-Startcodon und an ihrem 3'-Ende eine BamHI-Schnittstelle auf. Ein doppelsträngiges DNA-Molekül, das auf vorstehender Sequenz basiert, wird mit BamHI/BgIII geschnitten und in den mit BamHI-gespaltenen und dephosphorylierten Expressionsvektor pCDNA.3 (vgl. vorstehend) inseriert, wodurch der Expressionsvektor pCZPP erhalten wird.

Ferner wird der Expressionsvektor pTKTHBV2 (vgl. vorstehend) verwendet, um das Fragment nt-HBV 1861-2136 mittels PCR zu amplifizieren. Der forward-Primer umfaßt die folgende Sequenz: XXX GGA TCC ACT GTT CAA GCC TCC AAG CTG. Diese umfaßt eine BamHI-Schnittstelle gefolgt von der Sequenz ntHB 1861-1881. Der backward Primer umfaßt die folgende Sequenz: XXX GAA TTC TGG ATC TTC CAA ATT AAC ACC CAC CCA. Diese umfaßt eine EcoRI Schnittstelle gefolgt von der Sequenz ntHBV 2139-2116. In einer zweiten PCR wird das Fragment ntHBV 2140-2480 amplifiziert, das an seinem 5'-Ende mit der für das RGD-Motiv kodierenden Sequenz erweitert ist. Der forward Primer umfaßt die folgende Sequenz: XXX GAA TTC CGA GGC GAC GCG TCT AGA GAC CTA GTA GTC. Diese umfaßt eine EcoRI-Schnittstelle gefolgt von der für das RGD-Motiv kodierenden Sequenz und der Sequenz ntHBV2140-2161. Der backward Primer umfaßt die folgende Sequenz: XXX AAG CTT TCC CCA CCT TAT GAG TCC AAG. Diese umfaßt eine HindIII-Schnittstelle und die Sequenz ntHBV 2480-2460.

Erhaltene Fragmente beider PCRs werden mit EcoRI gespalten und

miteinander ligiert. Das Ligationsprodukt wird als Template für eine weitere PCR verwendet, wobei als forward Primer jener der ersten PCR und als backward Primer jener der zweiten PCR verwendet werden. Erhaltene PCR-Fragmente werden mit BamHI/-HindIII gespalten und in den mit BamHI/HindIII-gespaltenen und dephosphorylierten Vektor pCZPP inseriert, wodurch der Expressionsvektor pCZPPHBcRGC erhalten wird. Dieser Expressionsvektor kodiert für HBcAg, das N-terminal die ZPP-Sequenz und im Bereich der Aminosäuren 79-82 die RGD-Sequenz enthält.

Des weiteren werden etwa 0.8×10^6 HepG2-Zellen mittels Liopfektion mit 4µg eines für HBV-Polymerase kodierenden Expressionsvektors und mit 2 µg pCDN3 transfiziert. Es wird auf die vorstehende Beschreibung von Beispiel 1,(D) verwiesen. Es wird eine mit HepG2-HBV Pol bezeichnete Zellinie erhalten.

Etwa 0.8×10^6 Zellen der Zellinie HepG2-HBV Pol werden mit 3µg von pCZPPHBc RGC und 3µg von pCVPHBVeGFP (vgl. Beispiel 1,B) mittels Lipofektion transfiziert. Es wird auf vorstehende Beschreibung von Beispiel 1,(E) verwiesen. Es werden erfindungsgemäße Partikel in reiner Form erhalten.

Beispiel 3: Nachweis der Expression einer in erfindungsgemäßen Partikeln vorliegender Nukleinsäure in Fibroblasten

Etwa 1×10^9 erfindungsgemäße Partikel von Beispiel 1 (E) bzw. Beispiel 2, werden in 100µl 0,9 % Saline gelöst und in die Schwanzvene von balb/c Mäusen injiziert. 48 h nach Injektion wird der Soleus- und der Tibialis anterior Muskel isoliert und in einem "tissue tag" langsam eingefroren. Aus den eingefrorenen Präparaten werden Kryoschnitte angefertigt und diese unter einem Fluoreszenzmikroskop bei Blauanregung analysiert.

Es wird eine grüne Fluoreszenz in den Fibroblasten erhalten, was die Expression des "green fluorescent protein" zeigt.

Beispiel 4: Herstellung und Reinigung eines erfindungsgemäßen Fusionsproteins

Es wird das erfindungsgemäße Fusionsprotein von Fig. 1 hergestellt. Hierzu wird die DNA von Fig. 1 am 5'-Ende mit einem BglII-Linker und am 3'-Ende mit einem BglII-Linker versehen und mit den entsprechenden Restriktionsenzymen nachgespalten. Das erhaltene BglII/BglII-Fragment wird in den BamHI-gespaltenen Expressionsvektor pQE8 inseriert, so daß das Expressionsplasmid pQE8/LHBs erhalten wird. Ein solches kodiert für ein Fusionsprotein aus 6 Histidin-Resten (N-Terminuspartner) und dem erfindungsgemäßen Fusionsprotein von Fig. 1 (C-Terminuspartner). pQE-8/LHBs wird zur Transformation von E.coli SG 13009(vgl. Gottesman, S. et al., J. Bacteriol. 148, (1981), 265-273) verwendet. Die Bakterien werden in einem LB-Medium mit 100µg/ml Ampicillin und 25µg/ml Kanamycin kultiviert und 4 h mit 60µM Isopropyl-ß-D-Thiogalactopyranosid (IPTG) induziert. Durch Zugabe von 6 M Guanidinhydrochlorid wird eine Lyse der Bakterien erreicht, anschließend wird mit dem Lysat eine Chromatographie (Ni-NTA-Resin) in Gegenwart von 8 M Harnstoff entsprechend der Angaben des Herstellers (Qiagen) des Chromatographie-Materials durchgeführt. Das gebundene Fusionsprotein wird in einem Puffer mit pH 3,5 eluiert. Nach seiner Neutralisierung wird das Fusionsprotein einer 18 % SDS-Polyacrylamid-Gelelektrophorese unterworfen und mit Coomassie-Blau angefärbt (vgl. Thomas, J.O. und Kornberg, R.D., J.Mol.Biol. 149 (1975), 709-733).

Es zeigt sich, daß ein erfindungsgemäßes Fusionsprotein in hochreiner Form hergestellt werden kann.

Beispiel 5: Herstellung und Nachweis eines erfindungsgemäßen Antikörpers

Ein erfindungsgemäßes Fusionsprotein von Beispiel 4 wird einer 18 %igen SDS-Polyacrylamid-Gelelektrophorese unterzogen. Nach Anfärbung des Gels mit 4 M Natriumacetat wird eine 38 kD Bande

aus dem Gel herausgeschnitten und in Phosphat gepufferter Kochsalzlösung inkubiert. Gel-Stücke werden sedimentiert, bevor die Proteinkonzentration des Überstandes durch eine SDS-Polyacrylamid-Gelelektrophorese, der eine Coomassie-Blau-Färbung folgt, bestimmt wird. Mit dem Gel-gereinigten Fusionspolypeptid werden Tiere wie folgt immunisiert:

Immunisierungsprotokoll für polyklonale Antikörper im Kaninchen

Pro Immunisierung werden 35 µg Gel-gereinigtes Fusionsprotein in 0,7 ml PBS und 0,7 ml komplettem bzw. inkomplettem Freund's Adjuvans eingesetzt.

Tag 0: 1. Immunisierung (komplettes Freund's Adjuvans)
Tag 14: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 28: 3. Immunisierung (icFA)
Tag 56: 4. Immunisierung (icFA)
Tag 80: Ausbluten

Das Serum des Kaninchens wird im Immunoblot getestet. Hierzu wird ein erfindungsgemäßes Fusionsprotein von Beispiel 4 einer SDS-Polyacrylamid-Gelelektrophorese unterzogen und auf ein Nitrocellulosefilter übertragen (vgl. Khyse-Andersen, J., J. Biochem. Biophys. Meth. 10, (1984), 203-209). Die Western Blot-Analyse wurde wie in Bock, C.-T. et al., Virus Genes 8, (1994), 215-229, beschrieben, durchgeführt. Hierzu wird das Nitrocellulosefilter eine Stunde bei 37°C mit einem ersten Antikörper inkubiert. Dieser Antikörper ist das Serum des Kaninchens (1:10000 in PBS). Nach mehreren Waschschritten mit PBS wird das Nitrocellulosefilter mit einem zweiten Antikörper inkubiert. Dieser Antikörper ist ein mit alkalischer Phosphatase gekoppelter monoklonaler Ziege Anti-Kaninchen-IgG-Antikörper (Dianova) (1:5000) in PBS. Nach 30-minütiger Inkubation bei 37°C folgen mehrere Waschschritte mit PBS und anschließend die alkalische Phosphatase-Nachweisreaktion mit Entwicklerlösung (36µM 5' Bromo-4-chloro-3-indolylphosphat,

400µM Nitroblau-tetrazolium, 100mM Tris-HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$) bei Raumtemperatur, bis Banden sichtbar werden.

Es zeigt sich, daß erfindungsgemäße, polyklonale Antikörper hergestellt werden können.

Immunisierungsprotokoll für polyklonale Antikörper im Huhn

Pro Immunisierung werden 40µg Gel-gereinigtes Fusionsprotein in 0,8 ml PBS und 0,8 ml komplettem bzw. inkomplettem Freund's Adjuvans eingesetzt.

Tag 0. 1. Immunisierung (komplettes Freund's Adjuvans)
Tag 28: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 50: 3. Immunisierung (icFA)

Aus Eigelb werden Antikörper extrahiert und im Western Blot getestet. Es werden erfindungsgemäße, polyklonale Antikörper nachgewiesen.

Immunisierungsprotokoll für monoklonale Antikörper der Maus

Pro Immunisierung werden 12µg Gel-gereinigtes Fusionsprotein in 0,25 ml PBS und 0,25 ml komplettem bzw. inkomplettem Freund's Adjuvans eingesetzt; bei der 4. Immunisierung ist das Fusionsprotein in 0,5 ml (ohne Adjuvans) gelöst.

Tag 0. 1. Immunisierung (komplettes Freund's Adjuvans)
Tag 28: 2. Immunisierung (inkomplettes Freund's Adjuvans; icFA)
Tag 56: 3. Immunisierung (icFA)
Tag 84: 4. Immunisierung (PBS)
Tag 87: Fusion

Überstände von Hybridomen werden im Western Blot getestet. Erfindungsgemäße, monoklonale Antikörper werden nachgewiesen.

Patentansprüche

1. Partikel, umfassend:
   (a) eine Proteinhülle mit einem Fusionsprotein, das ein Virus-Protein, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle umfaßt, und
   (b) eine in der Proteinhülle vorliegende Nukleinsäure, die Sequenzen für ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist.

2. Partikel nach Anspruch 1, wobei das Virus-Protein von einem Adenovirus, Adeno-assoziierten Virus, Vacciniavirus, Baculovirus oder Hepadna-Virus stammt.

3. Partikel nach Anspruch 2, wobei das Hepadna-Virus ein Hepatitis B-Virus ist.

4. Partikel nach einem der Ansprüche 1-3, wobei das Virus-Protein ein Oberflächenprotein ist.

5. Partikel nach Anspruch 4, wobei das Oberflächenprotein ein LHBs ist.

6. Partikel nach einem der Ansprüche 1-3, wobei das Virus-Protein ein Core-Protein ist.

7. Partikel nach Anspruch 6, wobei das Core-Protein ein HBcAg ist.

8. Partikel nach einem der Ansprüche 1-7, wobei das Zellpermeabilität-vermittelnde Peptid die folgende Aminosäuresequenz aufweist:

9. Partikel nach einem der Ansprüche 1-8, wobei die heterologe zellspezifische Bindungsstelle RGD ist.

10. Partikel nach einem der Ansprüche 1-9, wobei das Fusionsprotein jenes von Fig. 1 oder 2 ist.

11. Verfahren zur Herstellung des Partikels nach Anspruch 1, wobei das Fusionsprotein ein LHBs und eine heterologe zellspezifische Bindungsstelle enthält, umfassend die folgenden Verfahrensschritte:

(a) Co-Transfektion von Zellen, die für ein Heptatitis B-Virus-Genom kodieren, wobei diese kein LHBs exprimieren, mit einem ersten Expressionsvektor, der für ein Fusionsprotein kodiert, das ein LHBs und eine heterologe zellspezifische Bindungsstelle umfaßt, und mit einem zweiten Expressionsvektor, der ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist, und

(b) Isolierung und Reinigung des Partikels.

12. Verfahren zur Herstellung des Partikels nach Anspruch 1, wobei das Fusionsprotein ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle aufweist, umfassend die folgenden Verfahrensschritte:

(a) Co-Transfektion von Zellen, die für eine HBV-Polymerase kodieren, mit einem ersten Expressionsvektor, der für ein Fusionsprotein kodiert, das ein HBcAg, ein Zellpermeabilität-vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle umfaßt, und mit einem zweiten Expressionsvektor, der ein Virus-spezifisches Verpackungssignal und ein Struktur-Gen aufweist, und

(b) Isolierung und Reinigung des Partikels.

13. Fusionsprotein, umfassend ein Virus-Protein, ein

Zellpermeabilität-vermittelndes Peptid und eine heterologe zellspezifische Bindungsstelle.

14. Fusionsprotein nach Anspruch 13, umfassend die Aminosäuresequenz von Fig. 1 oder 2 oder eine hiervon durch ein oder mehrere Aminosäuren unterschiedliche Aminosäuresequenz.

15. DNA, kodierend für das Fusionsprotein nach Anspruch 13.

16. DNA, kodierend für das Fusionsprotein nach Anspruch 14, umfassend:

(a) Die DNA von Fig. 1 oder 2 oder eine hiervon durch ein oder mehrere Basenpaare unterschiedliche DNA, oder

(b) eine mit der DNA von (a) über den degenerierten gentischen Code verwandte DNA.

17. Expressionsvektor, kodierend für die DNA nach Anspruch 16.

18. Verwendung des Partikels nach einem der Ansprüche 1-10 zur Gentherapie von Zellen und Geweben.

atgggccgtggcgaaggagctggagcattcgggctgggtttcaccccaccgcacggaggccttttg
gggtggagccctcaggctcagggcatactacaaactttgccagcaaatccgcctcctgcctccacc
aatcgccagacaggaaggcagcctaccccgctgtctccacctttgagaaacactcatcctcaggcc
atgcagtggaattccacaacctttcaccaaactctgcaagatcccagagtgagaggcctgtatttc
cctgctggtggctccagttcaggagcagtaaaccctgttccgactactgcctctcccttatcgtca
atcttctcgaggattggggaccctgcgctgaacatggagaacatcacatcaggattcctaggaccc
cttctcgtgttacaggcggggtttttcttgttgacaagaatcctcacaataccgcagagtctagac
tcgtggtggacttctctcaattttctaggggaactaccgtgtgtcttggccaaaattcgcagtcc
ccaacctccaatcactcaccaacctcctgtcctccaacttgtcctggttatcgctggatgtgtctg
cggcgttttatcatcttcctcttcatcctgctgctatgcctcatcttcttgttggttcttctggac
tatcaaggtatgttgcccgtttgtcctctaattccaggatcctcaaccaccagcacgggaccatgc
cgaacctgcatgactactgctcaaggaacctctatgtatccctcctgttgctgtaccaaaccttcg
gacggaaattgcacctgtattcccatcccatcatcctgggctttcggaaaattcctatgggagtgg
gcctcagcccgtttctcctggctcagtttactagtgccatttgttcagtggttcgtagggctttcc
cccactgtttggctttcagttatatggatgatgtggtattgggggccaagtctgtacagcatcttg
agtccctttttaccgctgttaccaattttcttttgtctttgggtatacatttaaacc

MGRGDGAGAFGLGFTPPHGGLLGWSPQAQGILETLPANPPPASTNRQSGRQPTPLSPPLRNTHPQA
MQWNSTTFHQTLQDPRVRGLYFPAGGSSSGTVNPVPTTVSPISSIFSRIGDPALNMENITSGFLGP
LLVLQAGFFLLTRILTIPQSLDSWWTSLNFLGGTTVCLGQNSQSPTSNHSPTSCPPTCPGYRWMCL
RRFIIFLFILLLCLIFLLVLLDYQGMLPVCPLIPGSSTTSTGPCRTCTTPAQGTSMYPSCCCTKPS
DGNCTCIPIPSSWAFGKFLWEWASARFSWLSLLVPFVQWFVGLSPTVWLSVIWMMWYWGPSLYSIL
SPFLPLLPIFFCLWVYI

FIG.1

atgcccatatcgtcaatcttctcgaggattggggaccctggatccactactgttcaagcctccaag
ctgtgccttgggtggctttggggcatggacatcgacccttataaagaatttggagctactgtggag
ttactctcgtttttgccttctgacttctttccttcagtacgagatcttctagataccgcctcagct
ctgtatcgggaagccttagagtctcctgagcattgttcacctcaccatactgcactcaggcaagca
attctttgctgggggggaactaatgactctagctacctgggtgggtgttaatttggaagatccagaa
ttccgaggcgacgcgtctagagacctagtagtcagttatgtcaacactaatatgggcctaaagttc
aggcaactcttgtggtttcacatttcttgtctcacttttggaagagaaaccgttatagagtatttg
gtgtctttcggagtgtggattcgcactcctccagcttatagaccaccaaatgcccctatcctatca
acacttccggaaactactgttgttagacgacgaggcaggtcccctagaagaagaactccctcgcct
cgcagacgaaggtctcaatcgccgcgtcgcagaagatctcaatctcgggaacctcaatgttagtat
tcc

MPLSSIFSRIGDPTVQASKLCLGWLWGMDIDPYKEFGATVELLSFLPSDFFPSVRDLLDTASALYR
EALESPEHCSPHHTALRQAILCWGELMTLATWVGVNLEDPEFRGDASRDLVVSYVNTNMGLKFRQL
LWFHISCLTFGRETVIEYLVSFGVWIRTPPAYRPPNAPILSTLPETTVVRRRGRSPRRRTPSPRRR
RSQSPRRRRSQSREPQC

Fig.2

SEQUENZPROTOKOLL

(1) ALLGEMEINE ANGABEN:

(i) ANMELDER:
(A) NAME: Dr. Eberhardt Hildt, Klinikum Rechts d. Isar
(B) STRASSE: Ismaninger Str.
(C) ORT: Muenchen
(E) LAND: Deutschland
(F) POSTLEITZAHL: 81675

(ii) BEZEICHNUNG DER ERFINDUNG: Partikel zur Gentherapie

(iii) ANZAHL DER SEQUENZEN: 19

(iv) COMPUTER-LESBARE FASSUNG:
(A) DATENTRÄGER: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) BETRIEBSSYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPA)

(v) DATEN DER JETZIGEN ANMELDUNG:
ANMELDENUMMER: DE 199 04 800.2

(2) ANGABEN ZU SEQ ID NO: 1:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 347 Aminosäuren
(B) ART: Aminosäure
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 1:

Met Gly Arg Gly Asp Gly Ala Gly Ala Phe Gly Leu Gly Phe Thr Pro
1 5 10 15

Pro His Gly Gly Leu Leu Gly Trp Ser Pro Gln Ala Gln Gly Ile Leu
20 25 30

Glu Thr Leu Pro Ala Asn Pro Pro Pro Ala Ser Thr Asn Arg Gln Ser
35 40 45

Gly Arg Gln Pro Thr Pro Leu Ser Pro Pro Leu Arg Asn Thr His Pro
50 55 60

Gln Ala Met Gln Trp Asn Ser Thr Thr Phe His Gln Thr Leu Gln Asp
65 70 75 80

Pro Arg Val Arg Gly Leu Tyr Phe Pro Ala Gly Gly Ser Ser Ser Gly
85 90 95

Thr Val Asn Pro Val Pro Thr Thr Val Ser Pro Ile Ser Ser Ile Phe
100 105 110

Ser Arg Ile Gly Asp Pro Ala Leu Asn Met Glu Asn Ile Thr Ser Gly
115 120 125

Phe Leu Gly Pro Leu Leu Val Leu Gln Ala Gly Phe Phe Leu Leu Thr
130 135 140

Arg Ile Leu Thr Ile Pro Gln Ser Leu Asp Ser Trp Trp Thr Ser Leu
145 150 155 160

Asn Phe Leu Gly Gly Thr Thr Val Cys Leu Gly Gln Asn Ser Gln Ser
165 170 175

Pro Thr Ser Asn His Ser Pro Thr Ser Cys Pro Pro Thr Cys Pro Gly
180 185 190

Tyr Arg Trp Met Cys Leu Arg Arg Phe Ile Ile Phe Leu Phe Ile Leu
195 200 205

Leu Leu Cys Leu Ile Phe Leu Leu Val Leu Leu Asp Tyr Gln Gly Met
210 215 220

Leu Pro Val Cys Pro Leu Ile Pro Gly Ser Ser Thr Thr Ser Thr Gly
225 230 235 240

Pro Cys Arg Thr Cys Thr Thr Pro Ala Gln Gly Thr Ser Met Tyr Pro
245 250 255

Ser Cys Cys Cys Thr Lys Pro Ser Asp Gly Asn Cys Thr Cys Ile Pro
260 265 270

Ile Pro Ser Ser Trp Ala Phe Gly Lys Phe Leu Trp Glu Trp Ala Ser
275 280 285

Ala Arg Phe Ser Trp Leu Ser Leu Leu Val Pro Phe Val Gln Trp Phe
290 295 300

Val Gly Leu Ser Pro Thr Val Trp Leu Ser Val Ile Trp Met Met Trp
305 310 315 320

Tyr Trp Gly Pro Ser Leu Tyr Ser Ile Leu Ser Pro Phe Leu Pro Leu
325 330 335

Leu Pro Ile Phe Phe Cys Leu Trp Val Tyr Ile
340 345

(2) ANGABEN ZU SEQ ID NO: 2:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 215 Aminosäuren
(B) ART: Aminosäure
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Peptid

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 2:

Met Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Thr Val Gln
1 5 10 15

Ala Ser Lys Leu Cys Leu Gly Trp Leu Trp Gly Met Asp Ile Asp Pro
20 25 30

Tyr Lys Glu Phe Gly Ala Thr Val Glu Leu Leu Ser Phe Leu Pro Ser
35 40 45

Asp Phe Phe Pro Ser Val Arg Asp Leu Leu Asp Thr Ala Ser Ala Leu
50 55 60

Tyr Arg Glu Ala Leu Glu Ser Pro Glu His Cys Ser Pro His His Thr
65 70 75 80

Ala Leu Arg Gln Ala Ile Leu Cys Trp Gly Glu Leu Met Thr Leu Ala
85 90 95

Thr Trp Val Gly Val Asn Leu Glu Asp Pro Glu Phe Arg Gly Asp Ala
100 105 110

Ser Arg Asp Leu Val Val Ser Tyr Val Asn Thr Asn Met Gly Leu Lys
115 120 125

Phe Arg Gln Leu Leu Trp Phe His Ile Ser Cys Leu Thr Phe Gly Arg
130 135 140

Glu Thr Val Ile Glu Tyr Leu Val Ser Phe Gly Val Trp Ile Arg Thr
145 150 155 160

Pro Pro Ala Tyr Arg Pro Pro Asn Ala Pro Ile Leu Ser Thr Leu Pro
165 170 175

Glu Thr Thr Val Val Arg Arg Arg Gly Arg Ser Pro Arg Arg Arg Thr
180 185 190

Pro Ser Pro Arg Arg Arg Arg Ser Gln Ser Pro Arg Arg Arg Arg Ser
195 200 205

Gln Ser Arg Glu Pro Gln Cys
210 215

(2) ANGABEN ZU SEQ ID NO: 3:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 663 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 3:

ATGCCCATAT CGTCAATCTT CTCGAGGATT GGGGACCCTG GATCCACTAC TGTTCAAGCC 60

TCCAAGCTGT GCCTTGGGTG GCTTTGGGGC ATGGACATCG ACCCTTATAA AGAATTTGGA 120

GCTACTGTGG AGTTACTCTC GTTTTTGCCT TCTGACTTCT TTCCTTCAGT ACGAGATCTT 180

CTAGATACCG CCTCAGCTCT GTATCGGGAA GCCTTAGAGT CTCCTGAGCA TTGTTCACCT 240

CACCATACTG CACTCAGGCA AGCAATTCTT TGCTGGGGGG AACTAATGAC TCTAGCTACC 300

TGGGTGGGTG TTAATTTGGA AGATCCAGAA TTCCGAGGCG ACGCGTCTAG AGACCTAGTA 360

GTCAGTTATG TCAACACTAA TATGGGCCTA AAGTTCAGGC AACTCTTGTG GTTTCACATT 420

TCTTGTCTCA CTTTTGGAAG AGAAACCGTT ATAGAGTATT TGGTGTCTTT CGGAGTGTGG 480

ATTCGCACTC CTCCAGCTTA TAGACCACCA AATGCCCCTA TCCTATCAAC ACTTCCGGAA 540

ACTACTGTTG TTAGACGACG AGGCAGGTCC CCTAGAAGAA GAACTCCCTC GCCTCGCAGA 600

CGAAGGTCTC AATCGCCGCG TCGCAGAAGA TCTCAATCTC GGGAACCTCA ATGTTAGTAT 660

TCC 663

(2) ANGABEN ZU SEQ ID NO: 4:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 1047 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: cDNA

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 4:

ATGGGCCGTG GCGAAGGAGC TGGAGCATTC GGGCTGGGTT TCACCCCACC GCACGGAGGC 60

CTTTTGGGGT GGAGCCCTCA GGCTCAGGGC ATACTACAAA CTTTGCCAGC AAATCCGCCT 120

CCTGCCTCCA CCAATCGCCA GACAGGAAGG CAGCCTACCC CGCTGTCTCC ACCTTTGAGA 180

AACACTCATC CTCAGGCCAT GCAGTGGAAT TCCACAACCT TTCACCAAAC TCTGCAAGAT 240

CCCAGAGTGA GAGGCCTGTA TTTCCCTGCT GGTGGCTCCA GTTCAGGAGC AGTAAACCCT 300

GTTCCGACTA CTGCCTCTCC CTTATCGTCA ATCTTCTCGA GGATTGGGGA CCCTGCGCTG 360

AACATGGAGA ACATCACATC AGGATTCCTA GGACCCCTTC TCGTGTTACA GGCGGGGTTT 420

TTCTTGTTGA CAAGAATCCT CACAATACCG CAGAGTCTAG ACTCGTGGTG GACTTCTCTC 480

AATTTTCTAG GGGGAACTAC CGTGTGTCTT GGCCAAAATT CGCAGTCCCC AACCTCCAAT 540

CACTCACCAA CCTCCTGTCC TCCAACTTGT CCTGGTTATC GCTGGATGTG TCTGCGGCGT 600

TTTATCATCT TCCTCTTCAT CCTGCTGCTA TGCCTCATCT TCTTGTTGGT TCTTCTGGAC 660

TATCAAGGTA TGTTGCCCGT TTGTCCTCTA ATTCCAGGAT CCTCAACCAC CAGCACGGGA 720

CCATGCCGAA CCTGCATGAC TACTGCTCAA GGAACCTCTA TGTATCCCTC CTGTTGCTGT 780

ACCAAACCTT CGGACGGAAA TTGCACCTGT ATTCCCATCC CATCATCCTG GGCTTTCGGA 840

AAATTCCTAT GGGAGTGGGC CTCAGCCCGT TTCTCCTGGC TCAGTTTACT AGTGCCATTT 900

GTTCAGTGGT TCGTAGGGCT TTCCCCCACT GTTTGGCTTT CAGTTATATG GATGATGTGG 960

TATTGGGGGC CAAGTCTGTA CAGCATCTTG AGTCCCTTTT TACCGCTGTT ACCAATTTTC 1020

TTTTGTCTTT GGGTATACAT TTAAACC 1047

(2) ANGABEN ZU SEQ ID NO: 5:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 35 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 5:

CCATATTCTT GGGAACAAGA TATCCAGCAC GGGGC 35

(2) ANGABEN ZU SEQ ID NO: 6:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 33 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 6:

GGATTGCTGG TGGAAGATAT CTGCCCCGTG CTG 33

(2) ANGABEN ZU SEQ ID NO: 7:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 33 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 7:

CAGCACGGGG CAGATATCTT CCACCAGCAA TCC 33

(2) ANGABEN ZU SEQ ID NO: 8:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 38 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 8:

GCCCCGTGCT GGATATCATC TTGTTCCCAA GAATATGG 38

(2) ANGABEN ZU SEQ ID NO: 9:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 36 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 9:

AAAAGATCTG GCCGTGGCGA AGGAGCTGGA GCATTC 36

(2) ANGABEN ZU SEQ ID NO: 10:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 30 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 10:

AAAAGATCTG GTTTAAATGT ATACCCAAAG 30

(2) ANGABEN ZU SEQ ID NO: 11:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 33 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 11:

CCCGATATCA TGTCATCTCT TGTTCATGTC CTA 33

(2) ANGABEN ZU SEQ ID NO: 12:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 30 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang

(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 12:

GGGGATATCG GTCGATGTCC ATGCCCCAAA 30

(2) ANGABEN ZU SEQ ID NO: 13:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 36 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 13:

GGGGGATCCC GATGTACGGG CCAGATATAC GCGTTG 36

(2) ANGABEN ZU SEQ ID NO: 14:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 27 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 14:

GGGGGATCCG CGGCCGCTTT ACTTGTA 27

(2) ANGABEN ZU SEQ ID NO: 15:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 57 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 15:

NNNAGATCTA TGCCCATATC GTCAATCTTC TCGAGGATTG GGGACCCTGG ATCCNNN 57

(2) ANGABEN ZU SEQ ID NO: 16:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 30 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 16:

NNNGGATCCA CTGTTCAAGC CTCCAAGCTG 30

(2) ANGABEN ZU SEQ ID NO: 17:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 36 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 17:

NNNGAATTCT GGATCTTCCA AATTAACACC CACCCA 36

(2) ANGABEN ZU SEQ ID NO: 18:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 39 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 18:

NNNGAATTCC GAGGCGACGC GTCTAGAGAC CTAGTAGTC 39

(2) ANGABEN ZU SEQ ID NO: 19:

(i) SEQUENZKENNZEICHEN:
(A) LÄNGE: 30 Basenpaare
(B) ART: Nucleotid
(C) STRANGFORM: Einzelstrang
(D) TOPOLOGIE: linear

(ii) ART DES MOLEKÜLS: Sonstige Nucleinsäure
(A) BESCHREIBUNG: /desc = "Primer"

(xi) SEQUENZBESCHREIBUNG: SEQ ID NO: 19:

NNNAAGCTTT CCCCACCTTA TGAGTCCAAG 30

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☑ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☑ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.